1982-03

A methodology for the evaluation of unit tactical proficiency at the National Training Center

Furman, John Scott; Wampler, Richard Lynn

Monterey, California. Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

A Methodology for the Evaluation of Unit
Tactical Proficiency
at the National Training Center

by

John Scott Furman

Richard Lynn Wampler

March 1982

Thesis Advisors: J. L. Ellis and S. H. Parry

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Methodology for the Evaluation of Unit Tactical Proficiency at the National Training Center		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis March 1982
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) John Scott Furman Richard LynnWampler		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE March 1982
		13. NUMBER OF PAGES 212
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Training; Training Analysis; Training Evaluation; ARTEP; Training Statistics; Normative Evaluation; National Training Center; NTC; Instrumented Battlefield; Training Readiness Profile; TRP; Training Measures of Effectiveness; MOE's; Combat Analysis; Combat Process.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The evaluation process currently planned for the National Training Center at Fort Irwin, California, is examined and a methodology proposed for evaluating unit tactical proficiency from the data accumulated. The concept of a Training Readiness Profile (TRP) is suggested as a concise method for assisting the Battalion Commander and his subordinates in meeting the training objectives of FM 71-1, FM 71-2, and ARTEP 71-2. This concept is

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

applicable to Armored and Infantry battalions and to their training programs as currently specified under Department of the Army Doctrine. This methodology is compatible with the automated information retrieval systems currently being specified for installation at the National Training Center.

A Methodology for the Evaluation of Unit Tactical Proficiency at the National Training Center

by

John S. Furman
Captain, United States Army
B.S., United States Military Academy, 1975

and

Richard L. Wampler
Captain, United States Army
B.S., United States Military Academy, 1972

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March, 1982

ABSTRACT

The evaluation process currently planned for the National Training Center at Fort Irwin, California, is examined and a methodology proposed for evaluating unit tactical proficiency from the data accumulated. The concept of a Training Readiness Profile (TRP) is suggested as a concise method for assisting the Battalion Commander and his subordinates in meeting the training objectives of FM 71-1, FM 71-2, and ARTEP 71-2. This concept is applicable to Armored and Infantry battalions and to their training programs as currently specified under Department of the Army doctrine. This methodology is compatible with the automated information retrieval systems currently being specified for installation at the National Training Center.

TABLE OF CONTENTS

LIST OF FIGURES

I. THE EVALUATION OF TRAINING FOR GROUND COMBAT

The mission of the United States Army is to fight in ground combat and win. From FM 100-5, Operations [Ref. 1],

> "The Army's primary objective is to win the land battle--to fight and win in battles, large and small, against whatever foe, wherever we may be sent to war." [Ref. 1:pg. 1-1]

This mission stands essentially unchanged since the earliest days of the Republic, yet the world has changed dramatically in the intervening years. Increasingly complex technology has expanded the range and lethality of all weapons on the battlefield to such an extent that:

> "We can expect very high losses to occur in short periods of time. Entire forces could be desroyed quickly if they are improperly employed..." (emphasis added) [Ref. 1:pg. 1-1]

Therefore, now as never before, the Army must train in peacetime to be victorious without the traditional long period of mobilization which has characterized the entry of the U.S. into all its other wars.

> "Today the U.S. Army must, above all else, prepare to win the first battle of the next war." [Ref. 1:pg. 1-1]

Within this century, training in the United States Army has been centered around the mobilization models first employed in WWI by the newly formed U.S. General Staff. This traditional model assumed that in the conduct of war, a

long period of time would be availible to raise, equip, and train an army while the continental U.S. remained protected by the formidable barriers of the Atlantic and Pacific oceans. This training followed a predictable cycle; the small standing Army formed a cadre-nucleus around which units were constructed from a large pool of conscripts. Training would begin with the individual combat skills, progress to squad, platoon, and company skills, and then these units would be aggregated to form regiments, divisions, and higher echelons, which proceeded through their own cycle of training and exercise. At the completion of this process, units were tested, and if judged fit, deployed to combat theaters.

This cycle worked admirably in WWII and enabled the U.S. to create a large and efficient war machine quickly. As this was a proven system, it formed the basis of all Army training until the early 1970's. At that point in time, in its final form, Army training consisted of the cycle as depicted in Figure 1.

Each stage was defined as listed below and the contents of each segment specified by an Army Training Program (ATP) which outlined the subjects to be taught in training, the number of hours to be spent in training, and the applicable

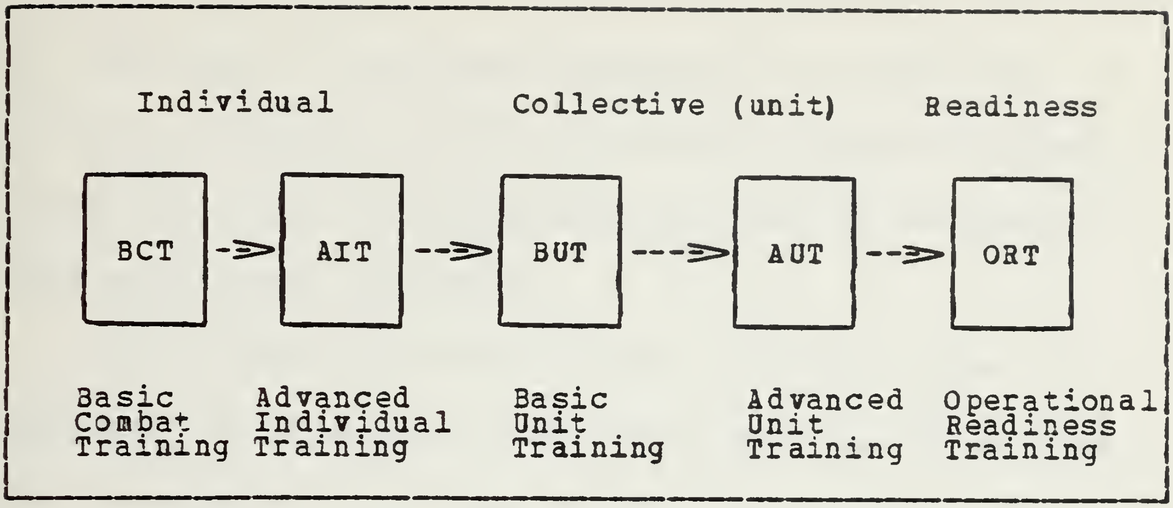


Figure 1: The Army Training Cycle (Obsolete)

resources and references to be used. The completion of a stage was evaluated by a graded test.

> BCT: Basic Combat Training- This training instilled the fundamentals of infantry combat to newly accessed personnel. [Ref. 2:pg. 19]

> AIT: Advanced Individual Training- Usually conducted in a formal setting at an institution which resulted in the award of a Military Occupation Specialty (MOS). [Ref. 2:pg. 19]

At this stage an individual was assigned to a unit. Here the cycle began again with:

> BUT: Basic Unit Training- The initial unit training designed to insure that squads, platoons, and Company/Battery/Troops can accomplish their TO&E missions. This phase, and all others, culminated in an Army Training Test (ATT) [Ref. 2:pg. 19].

> AUT: Advanced Unit Training- Training designed to insure that Battalions and Brigades can accomplish their TO&E missions. This phase also culminated in an ATT. [Ref. 2:pg. 19]

Finally, units progressed to the maneuver phase where joint combat arms maneuver characterized the training. This was:

ORT: Operational Readiness Training- Training designed to maintain the highest possible state of combat proficiency. [Ref. 2:pg. 21]

At this point, a unit was considered to be deployable and trained. Testing by ORTT (Operational Readiness Training Test) was regularly accomplished in order to determine if this status was maintained.

This system was fatally flawed, however. Under the pressures of personnel turbulence created by the Vietnam War, standards of training could not be maintained in an orderly cycle and unit readiness suffered greatly. When this deficiency was coupled with the need to maintain standing U.S. forces at peak readiness all of the time, and not just for a short period of time following the completion of a training cycle, it became clear that a reform of the Army training structure was necessary. In order to affect this reform,the Board for Dynamic Training was constituted in 1971 at the direction of the the Chief of Staff of the Army (CSA) and reported its conclusions on 17 Dec 1971 to the CSA regarding Army training. From these conclusions the concept of performance-oriented training was born. [Ref. 3]

Performance-Oriented Training (POT) was designed to do away with the old mobilization cycle of training. In it, a soldier was required to learn by doing a task under a given

set of conditions to a specified standard. The previous ATP had specified only hours of exposure to training which, as any noncommissioned officer (NCO) knows, does not guarantee the absorption of the skill being taught by the soldier. Although "hands-on training" by doing had been the teaching policy for many years in the Army, this new method emphasized the attainment of a standard as the final goal, as opposed to the simple completion of the requisite number of hours of training with or without acquiring the desired skill. Under the POT program training would continue (by doing) until the desired standard was achieved. This concept was such a dramatic departure from previous Army practice that the program took several years to introduce to the field.

This philosophy was refined and promulgated to the field Army through FM 21-6, How to Prepare and Conduct Military Training, 3 Nov 1975 [Ref. 4]. FM 21-6 officially made POT Army doctrine. Trainers were defined as those:

> "Whose duties include the requirement to prepare, conduct, and evaluate training..." [Ref. 4:pg. 2]

and these trainers were instructed that:

> "The last element of the training objective which you must develop is the training standard. These are needed to insure that the soldiers undergoing training will be able to perform the commander's objectives... Training standards are normally expressed in terms of measurement... or in terms of specific procedures which must be followed..." [Ref. 4:pg. 13]

Thus, the key to the training system as it stands today in the Army is the training standard. The initial question then facing the unit trainer is how well does the soldier, or collectively his unit, need to be trained? Beyond this question of basically how good is good enough, lies the yet more difficult question: how do you evaluate the <u>accomplishment</u> of the training standard, particularly for large units? It has been shown that setting the standard is difficult enough; how do you determine if men, units, and weapons are capable of accomplishing it? For larger elements, given the intangible nature of war, this question becomes a crucial issue, and one very difficult to answer.

A. THE PROBLEM: THE EVALUATION OF TRAINING STANDARDS FOR LARGE UNITS

With the dramatic shift from the accomplishment of training programs to the accomplishment of training standards, the Army incurred a new burden. Previously an ATP (or any cycle of an ATP) could be evaluated on a graded basis by administering an ATT to the unit. This could no longer suffice as an evaluation technique, for the prescribed training objective could only be considered reached when the unit or individual could perform the specified task to the given standard. The reliance of a

unit on a repetitive cycle of "Train-Test" in lockstep fashion was broken. In its place, a new system for training management was prescribed. In the introduction of this system the Army specified a cycle of "Train-Evaluate-Train" which encompassed the new philosophy that training in general should be an iterative learning experience. TC 21-5-7, Training Management in Battalions [Ref. 5], was issued in December 1977 to introduce the new cycle (see Figure 2).

This general concept of training required action by two agencies: the U.S. Army Training and Doctrine Command (TRADOC) and the training combat unit. TRADOC provided the defined tasks, conditions and standards to the units through the medium of the Army Training and Evaluation Program (ARTEP) [Ref. 6] and with the ARTEP as a guide it became a unit responsibility to complete the iterative "Train-Evaluate-Train" cycle. This concept also applied to individual, collective, and unit training (see Figure 3).

However, moves taken to decentralize training in the early 1970's [Ref. 2:pg. 4] effectively devolved this cyclic responsibility entirely to the battalion level. TC 21-5-7 states bluntly:

> "The clear intent of the policy of decentralization is to fix such responsibilities at battalion level, for the

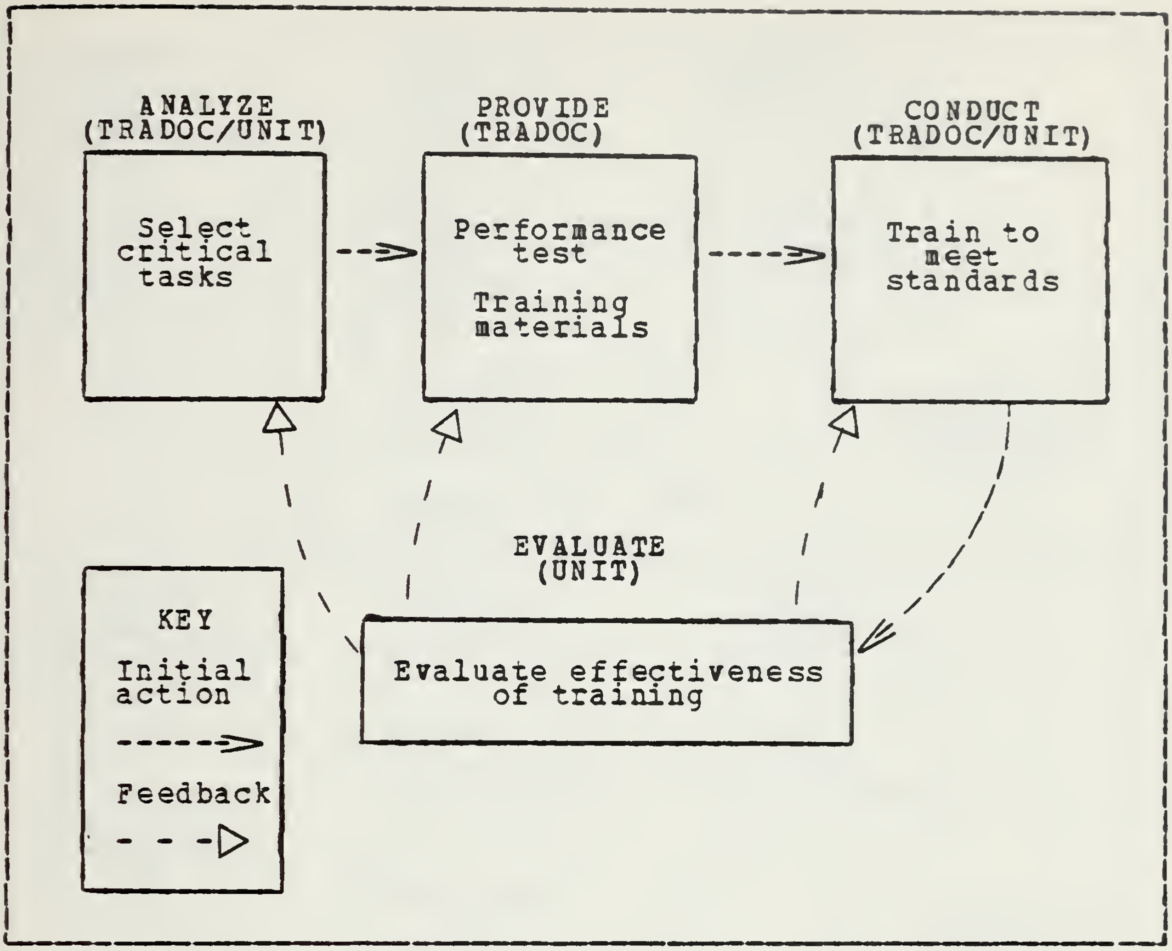


Figure 2: The General Concept of Training

battalion is the lowest echelon at which there is a staff to assist with such (training) duties." [Ref. 5:pg. 20]

Those duties were multitudinous, but most importantly, they stated that the battalion headquarters: "Assigns tasks for training and evaluates results." [Ref. 5:pg. 20]

Thus, unless carefully monitored through external ARTEP evaluations, the battalion became a training entity subject to its own set of training standards. In theory, external ARTEPs are scheduled and accomplished by a brigade

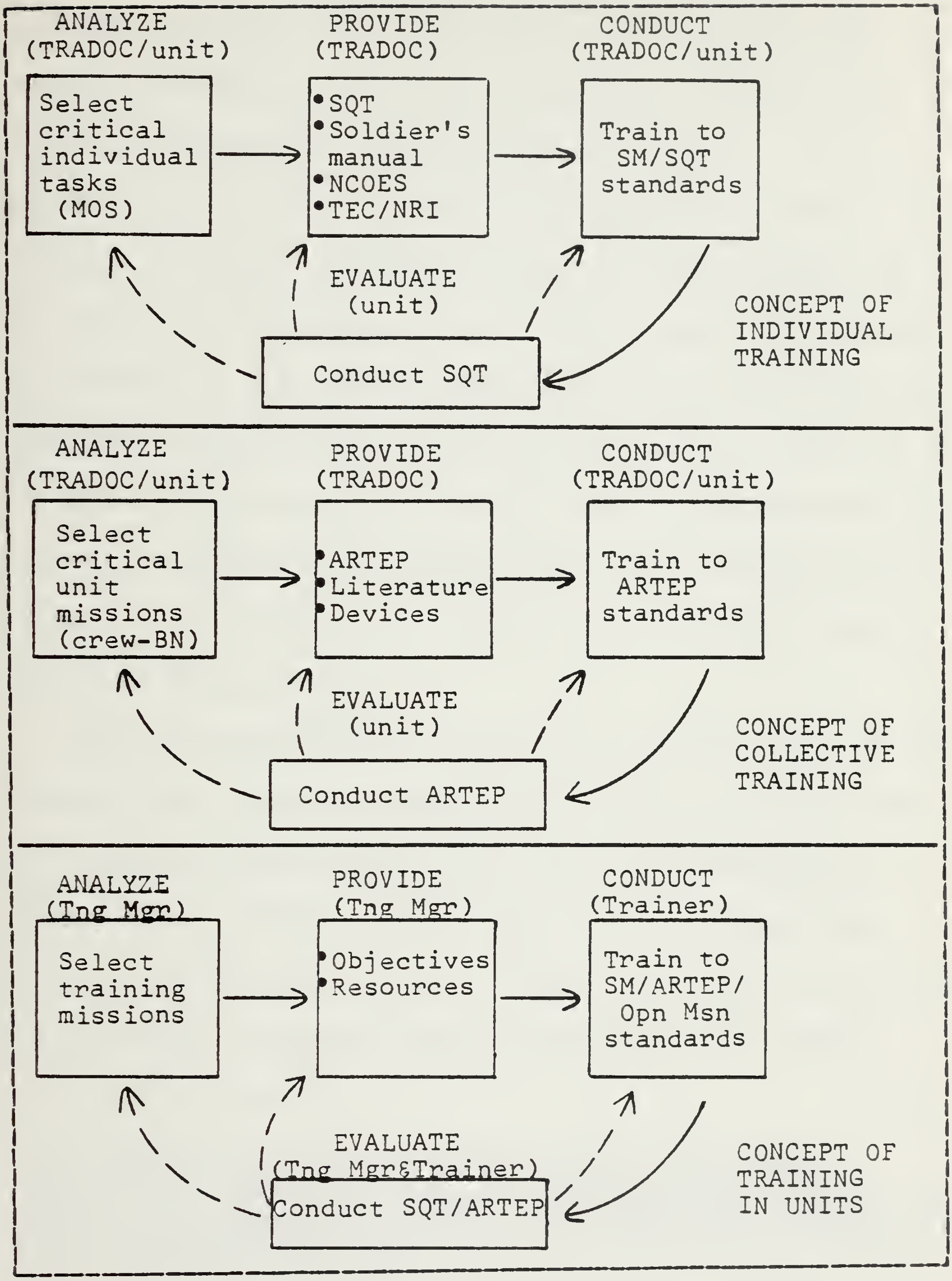


Figure 3: The Concept of Training at Levels

headquarters or higher echelon. Due to the number of evaluator personnel needed to administer an ARTEP, such events are usually scheduled by a division which employs another battalion's command group as evaluators. This perpetuates a vicious cycle; if one battalion is below standard, it cannot evaluate another battalion beyond its own level of competence. Hence the Army wide standard once envisioned for all units becomes less and less attainable as evaluation skills remain linked to the training level of other sister battalions. It thus becomes quite easy for two armored divisions, as an example, to report the same level of training readiness in all honesty and yet have totally disparate capabilities.

This situation becomes all the more critical when it is realized that the ARTEP tasks and standards are vague, and are frequently interpreted differently by individuals of varying skill levels. As an example, for a combat-ready tank and mechanized infantry battalion task force, (Level 1) the evaluation standards for the performance of a hasty attack are as follows:

> "b. Coordination of artillery, mortars, air defense, tactical air, engineers, and attack helicopters will support the scheme of maneuver. Maneuver units respond to task force commander's orders using overwatch positions, maximum suppressive fires, and the terrain for protection from opposing force fire.

c. Task force maintains the momentum of the attack and retains the initiative, aggressively seeking the opposing force's weakness. Suppressive fires will be coordinated to neutralize the opposing force's ability to react. Task force assets will be concentrated to overwhelm the opposing force at selected locations and control key terrain features.

d. Objective is secured within a reasonable time without sustaining excessive personnel and equipment losses. (Evaluator judgment.)" [Ref. 6:pg. 8-2-2]

The use of such phrases as "coordination ...will support the scheme of maneuver" and "task force maintains (the) momentum of the attack and retains the initiative" is not conducive to exacting evaluation. Such words can mean many things to many people. Even scholars of war and veterans of combat can argue justifiably over the definition of "retaining the initiative". Such descriptors are, in many cases, unintelligible to junior officers and NCOs with no combat experience. Yet these individuals are most often directly involved with the evaluation of sister elements in the accomplishment of an ARTEP.

Clearly then, although the concept of "Train-Evaluate-Train" is sound in theory, it lacks in the execution of the required evaluation. The problem, indeed, is two-fold: the establishment of accurate and concise training standards is extremely difficult to do in practice; but even more seriously, the evaluation of such standards by the untrained eye of an observer whose own experience and training may be

no better (and indeed, may be worse) than that of the individuals being evaluated provides little tangible information about the unit's real combat capability. Some current training programs are directed toward alleviating this deficiency; one in particular, the National Training Center at Ft. Irwin, California, may hold the key for breaking the vicious cycle of half-competent evaluation.

B. CURRENT TRAINING EVALUATION METHODS

The methods used to evaluate training accomplishments are fairly limited today. Many of them date back centuries to the simple concept of drill; others involve laser technology and fast computers. The training systems available to unit commanders at the present fall into three basic categories and possess varying degrees of exactness in evaluation. They can be classified from highest to lowest in terms of evaluating capability as the Skill Qualification Test (SQT)-Crew Drill-ARTEP hierarchy, the Multiple Integrated Laser Engagement Sysytem (MILES) training device (including all such variants of a hit/kill system) and various field maneuvers, such as the Field Training Exercise (FTX), the Command Post Exercise (CPX), and the Tactical Exercise Without Troops (TEWT).

1. <u>The SQT-Crew Drill-ARTEP Hierarchy</u>

Most organizations are constructed of vertical hierarchies and the Army is no exception to this rule. With the introduction of POT and the destruction of the old training cycle, certain vacuums were formed in the hierarchial progression of training from individual to unit. These gaps have now been filled by what can loosely be

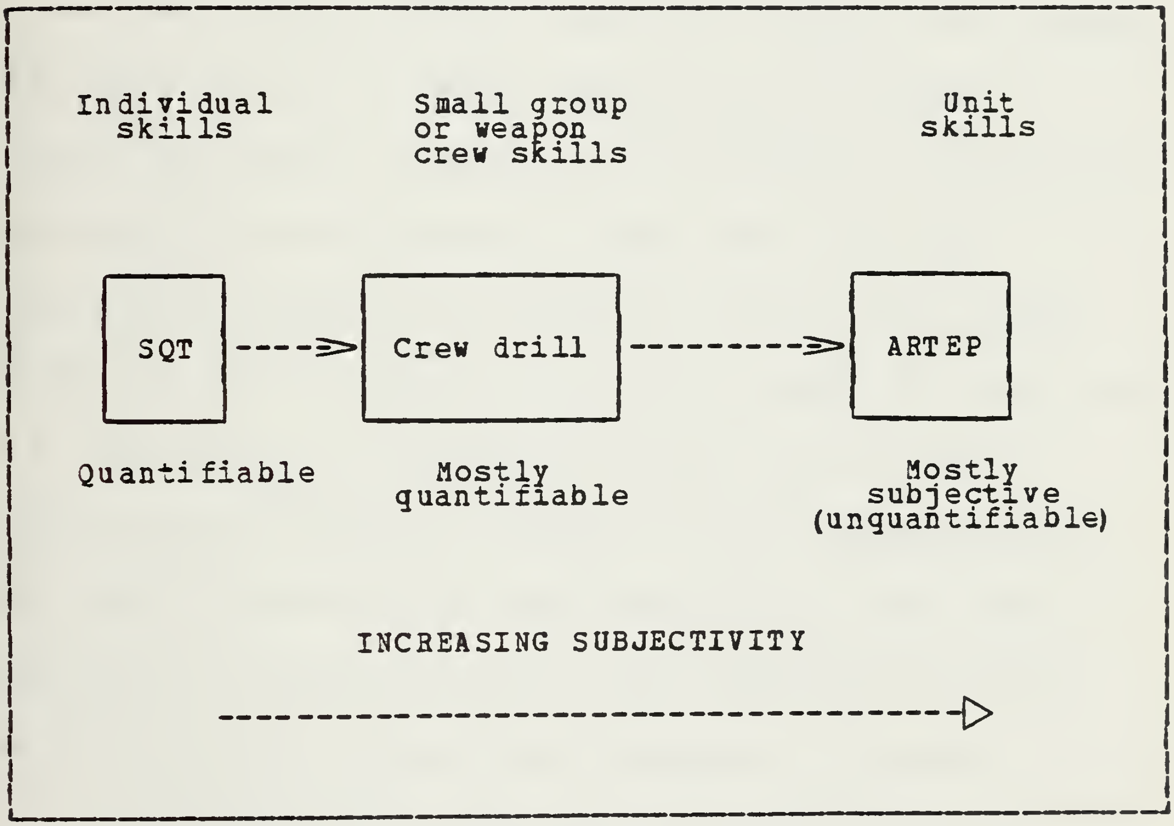


Figure 4: The SQT-Crew Drill-ARTEP Hierarchy

termed the SQT-Crew Drill-ARTEP hierarchy (see Figure 4). With respect to individual training, the Army has published the Soldier's Manual [Ref. 7] as a source document for the

Skill Qualification Test, which is a POT version of the old MOS (Military Occupation Specialty) Test designed to evaluate individual skills. This test is reasonably well structured, that is, the performance standards are quantifiable in terms of a go/no go performance. Thus, the SQT evaluates individual training by asking the soldier to perform a series of unambiguous mostly physical tasks which are graded either pass or fail.[1]

In the intermediate range of training crews, squads, and small collective units, the Army has no formal evaluation system, although the Army Research Institute for the Behaviorial and Social Sciences (ARI) has been tentatively tasked with arranging a "squad battle drill" program.[2] Several informal evaluation systems do exist which are fairly quantitative despite their lack of formal institutional standing. Examples of such evaluations are tank gunnery evaluations (Tank Table VIII) which require a quantitative passing score for crew performance; infantry machine-gun crew qualification standards, artillery gun crew

---

[1]It appears at the time of this writing that the Congress will refuse further funding for the SQT program. What impact this will have upon future individual training evaluation is unclear at this time.

[2]Based on a conversation with Dr. J. Banks of ARI, June 1981. The tasking is to develop a series of standardized exercises for small units, initially infantry squads, to formalize their training in accordance with the Train-Evaluate-Train cycle.

shooting standards, and the like. All of these evaluations are slightly less quantitative than the SQT (there is no way to evaluate the driver's shifting skill, for example, on Tank Table VIII by other than the fact that the tank proceeded downrange from target to target) but nonetheless, all evaluations of this sort result in a quantitative score against which a standard can be adjudged.

The final segment of this heirarchy is the previously discussed ARTEP. The ARTEP itself is segmented into platoon, company, and battalion portions, most of which rely on a high degree of subjectivity in their assessments. The inherent weaknesses of the ARTEP have already been discussed in the previous section, and generally these weaknesses apply to all echelons of the ARTEP with some degree of increasing subjectivity and ambiguity as the unit grows larger. This concept of a hierarchy is summarized by Figure 4, showing the direction of increasing subjectivity in evaluation.

2. CPX, FTX, TEWT, and Other Field Maneuver

This family of exercises, which consist mostly of field maneuver without an explicit opponent can be evaluative or educational. Command Post Exercises (CPX) and Tactical Exercises Without Troops (TEWT) are intended to

train staffs and commanders without the expense of placing troops in the field. Such exercises are beneficial to the leaders involved, but when these maneuvers are made into evaluations they are almost wholly subjective and, in fact, are missing the few benefits of the vague ARTEP system in that they totally lack established tasks, conditions, and standards. They also suffer from an additional weakness of possessing fewer observers than the normal ARTEP. Thus, for evaluative purposes, the CPX and TEWT offer less value than the ARTEP.

The field training exercise (FTX) suffers from the same inherent weakness when used as an evaluation. FTX's are quite often "canned" with little or no free play, and provide scant opportunity for any worthwhile evaluation of a unit's combat capabilities. Most often the FTX is used as a tool for "exercising" the assets of a division by forcing the service support units to operate in the field for several days. FTX's in their structured, scripted form offer very little training to combat elements and cannot be considered to be efficient or cost effective as evaluations of training.

3. <u>The MILES Training Device</u>

The MILES (Multiple Integrated Laser Engagement System) training device is not an evaluation methodology in itself but rather is a tool to be used in conjunction with other forms of training to increase realism and thus enhance evaluation. MILES consists of varying types of eye-safe lasers which can be fitted to most Army direct fire weapons from the M16 rifle to the 105mm tank gun. These devices are sound actuated, requiring the soldier to possess a blank round of ammunition and a functioning weapon in order to trigger the laser. This forces units to accurately perform maintenance and supply functions in real time. Each player in a MILES exercise is then fitted with sensors which can detect the strike of another MILES laser beam on the individual item, be it a single foot soldier or a tank, truck, etc. If a kill is registered, it is signified in various ways. Should an infantryman be "shot", for example, an audible tone sounds to indicate that he is a casualty and his MILES weapon is rendered inoperative. For a larger item such as a tank or APC, the kill indicator is appropriately larger; a smoke grenade is set off and a flashing signal started in addition to the instrumented weapon being rendered inoperative. A "dead" MILES system can only be

reactivated by a controller key device. In this manner the system can inflict realistic firepower casualties upon a player unit, and so represent combat losses. Weapon lasers are also coded internally within the beam to reflect weapon type. As an example of this it would be impossible to kill a tank with the beam of an M16 coded laser, while a TOW coded beam could inflict such a casualty. The MILES supporting software also contains a mini-computer which randomly induces a miss based on the actual probability of kill of the firing weapon. Thus, although it is easier to aim the laser and strike a sensor than it is to strike the target with a real round, the system automatically induces realistic weapon performance to simulate the effect of real fire from the given weapon.

MILES promises to add a degree of realism to training exercises such as has never before been known. Having the ability to "kill" an opponent greatly increases the desire of troops to wholeheartedly participate in mock battles, while possessing the ability to be "killed" causes them to seek cover more effectively and learn more quickly the price of their mistakes.[3] Therefore, the MILES system

---

[3]For an excellent discussion of the learning rate increase attributable to hit/kill systems see the REALTRAIN validation studies of the ARI (References 8,9,10, and 11).

can be seen as a multiplier of learning effects in training, but contains only an indirect methodology for evaluation. When utilized within the framework of an ARTEP style evaluation, it dramatically increases the degree of objectivity in the evaluation by removing the need for a subjective guess at casualties during a force-on-force exercise.

C. THE SHORTFALL: THE NEEDS OF EVALUATION VERSUS THE CONSTRAINTS OF TRAINING

In considering the weaknesses of the current Army training systems it becomes obvious that there is a dramatic shortfall between the the theory of the train-evaluate-train cycle and the hard realities of current training constraints. The lack of dedicated evaluator personnel forces the use of sister unit peers as pro tempore evaluators; the lack of a dedicated opposing force (OPFOR) utilizing threat tactics results in sister battalions providing the aggressors in training, causing US units to train against US tactics; and finally, the vagueness of the tasks and standards in the current ARTEP's of large combat units effectively eliminates the chance of efficient evaluation.

These shortfalls have not gone unnoticed in the Army training community. ARI was tasked in 1976 to evaluate the REALTRAIN system, a precursor to the MILES which utilized only optical spottings to provide "kill" information. In 1977 and 1978, tests were conducted to study the learning effect of this system upon troops in training. In constructing these tests, Scott, Meliza, and others effectively pinpointed the necessary elements to be added to the training system in order to obtain statistically reliable information regarding learning effect. [Refs. 8,9,10,11]

Their methodology for eliminating the subjectivity of traditional Army training methods consisted of a four-fold approach: better determination of terminal mission outcome (mission accomplishment) through the use of hit/kill systems (REALTRAIN or MILES); the elimination of the vague task descriptions in the ARTEP through the determination of intermediate subtasks (top-down analysis of combat processes); the use of dedicated OPFOR to provide realistic aggressors; and the use of event-specific trained evaluators to insure standardized evaluations. [Ref. 8,p.1-4] This methodology will be examined in detail.

1. Improved Determination of Mission Outcome

The lack of an objective method for determining mission accomplishment in an ARTEP is probably the program's biggest weakness. In testing the effect of the REALTRAIN system, Scott, Meliza, et al., agreed that:

> "The ARTEP initially suffered from some critical weaknesses; one was the inability to objectively determine tactical mission outcomes. However,the introduction of tactical engagement simulation training methods such as SCOPES, REALTRAIN, and eventually, the Multiple Integrated Laser Engagement System (MILES) has alleviated this problem. Using these methods, commanders have the capability to conduct two-sided, free-play tactical exercises with credible casualty assessment and weapons signature effects, and a high degree of realism." [Ref. 8:pg. 1]

This problem can then, for most practical purposes be considered to be solved if units engaged in training utilized the MILES system. The only major weakness remaining in the representation of weapons effects is in the play of indirect fire, as the effect of artillery upon troops during training still cannot be adequately simulated. For all other purposes, assuming all combatant elements possess the system during training, MILES will effectively determine mission outcome just as real battle would by demonstrating who-killed-who in combat.

2. Elimination of Vague Tasks Through Subtask Analysis

The second major obstacle to effective evaluation proved to be lack of detail concerning the combat processes:

"A second major weakness in the ARTEP is a general superficiality, which results in inadequate guidance for Army trainers...If a unit fails to accomplish its mission, this general guidance gives the trainer no help in determining reasons for failure." [Ref. 8:pg. 1]

In order to measure task performance accurately more concrete, identifiable subtasks had to be generated and measured. These were defined in the following way:

"Consistent with the performance-oriented, criterion-referenced approach, critical intermediary tasks are defined as those that substantially increase the probability of mission accomplishment." [Ref. 8:pg. 2]

The steps taken to identify these tasks were:

"(a) identification of candidate intermediate tasks;

(b) development of objective measures of proficiency for these tasks;

(c) determination of the test conditions necessary to gather these data reliably;

(d) correlation of intermediate task proficiency with mission outcomes;

(e) identification of those tasks which correlate most highly with mission outcomes." [Ref. 8:pg. 2]

This procedure essentially constituted a top-down analysis of the given mission, in this case a squad movement to contact (this test was repeated later for a tank platoon). The key point here is that once sufficient detail regarding the combat processes was established, measures of effectiveness for each subtask could be devised (see Figure 5). Thus, accurate identical assessments of performance could then be attained for each replication. This forms the basis for constructing statistically reliable data; as much subjectivity as possible has been eliminated from the ARTEP

AT PHASE LINE "A"

17. Call NCS when point crosses phase line "A".	____ (check)
18. From phase line "A" to potential threat, how many times did the element leader communicate with the squad?	____ (number) (x)
19. Are 3 or more men in your element within 10 meters of one another?	____ (number) (x)
20. Number of men in your element with which your element leader has visual contact. (Enter "x" if you could not observe.)	____ (number) (x)
21. How far is point ahead of squad? (Enter "x" if not observed.)	____ (meters) (x)
22. Was the element leader part of the point?	Yes (2) No (1) N/Obs (0)
23. Is point man covered by at least one man?	Yes (2) No (1) N/Obs (0)
24. Number of men in your element moving in the open. (Enter "x" if not observed.)	____ (number) (x)

Figure 5: An Example of Subtask MOE's for a Squad Mission

evaluation by the use of subtasks. (Dr. Scott, et al, eventually analyzed this data using Tukey's HSD* test and constructed standard Analysis of Variance (ANOVA) tables to support their conclusions [Ref. 8:pp. 153-154].) The value of this procedure did not escape the personnel at ARI;

> "The payoff from this research should be empirical identification and validation of critical intermediate

---

*Tukey's Honest Significant Difference Test (sometimes called the w procedure). This is a multiple comparison test. It is designed for making all pairwise comparisons among a set of means, in order to determine which differences are significant.

tasks as well as determination of objective measures. These measures of task proficiency and measurement procedures can be incorporated into the ARTEP."[Ref. 8:pg. 2]

3. The Use of a Dedicated OPFOR

In order to provide the necessary consistency of response to the tested units (squad, platoon) a dedicated OPFOR was used in each test to insure the opposition would be reasonably constant. This gave each tested unit the same experience and allowed the measurement of learning effect between REALTRAIN-trained units and traditionally (ARTEP only) trained units. This procedure would also allow, on a large scale such as in battalion vs. battalion encounters, the application of appropriate OPFOR tactics to training units. This effect is significant, for in deciding to establish the NTC the Army has relied heavily on Navy and Air Force experience with the use of trained OPFOR:

> "Moreover, it became apparent that training for air-to-air combat, which pitted TAC flyers against other U.S. pilots using U.S. tactics in identical aircraft, was not useful training for dogfights with North Vietnamese MIG's. The answer, TAC discovered, was to establish squadrons equipped with aircraft which resembled the MIG in size and operating capability, manned by pilots trained in Soviet-style tactics. The U.S. Navy had used this technique since 1969, and its pilots so trained, out-performed TAC sixfold in ratio of kills to losses." [Ref. 12:pg. 3]

The necessity for utilizing a dedicated OPFOR employing enemy tactics is obvious if it is desired to duplicate the "learning effect" of combat (see Figure 6).

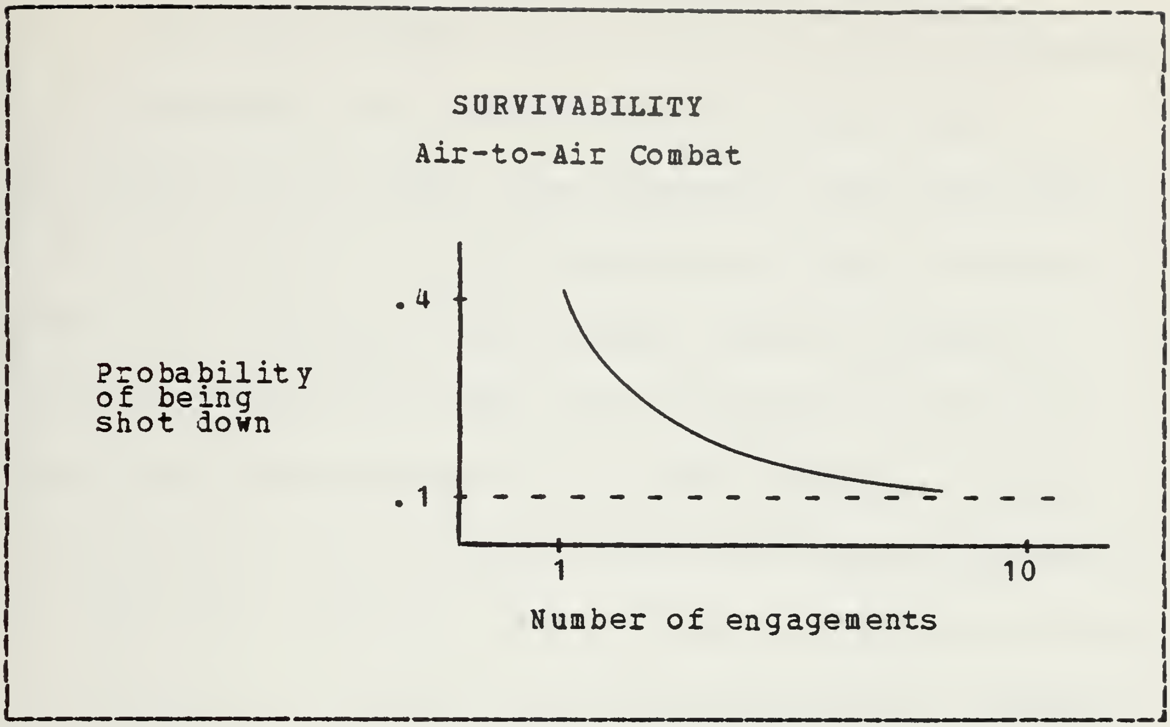


Figure 6: The Learning Effect of Battle

> "The battle statistics indicate to TAC that combat is a powerful trainer--whereas American pilots in their first combat engagement have had only a 40 percent chance of surviving, by their tenth engagement their chances of winning had increased to 90 percent." [Ref. 12:pg. 2]

Therefore, the use of special OPFOR troops in evaluating training for US units is not a luxury, but rather a necessity: it duplicates, as nearly as possible, the learning effect of combat and allows for the replication of training exercises so that training progress can be statistically measured.

4. <u>The Use of Trained Professional Evaluators</u>

The major reason for the use of trained evaluators in the REALTRAIN tests was to insure that all test procedures and scenarios were accurately replicated [Ref. 8:pg. 7]. The use of such evaluators at least insures a degree of consistency. On a larger scale, the use of trained evaluators in ARTEP type exercises for battalion sized units would accomplish the same thing, i.e. insure replicability, but also prevent the other major weakness of the ARTEP from appearing: peers would not evaluate elements of like skill or ability. With time and practice through training, even a junior evaluator could accurately report on the specific tasks for which he is trained, thus eliminating the blind-leading (evaluating)-the-blind effect common to the usual ARTEP. The degree of evaluation possible under this system is a great deal more incisive than that of the traditional ARTEP method.

The end result of this four-fold approach to improving evaluation is the shortfall: the amount of effort needed to bridge the gap between what is currently available due to the constraints of training, and what is necessary to produce a statistically reliable evaluation of training. These shortfalls have been bridged by the creation of the

National Training Center (NTC), for all the elements indigenous to the approach used by ARI are contained in the structure of training and evaluation conducted at the NTC. This is an exciting fact, for it establishes a statistically reliable training evaluation for the first time on a large-unit scale. A background disscussion of the NTC's technical capabilities follows, in order that its capacity for significantly improved evaluation may be better understood.

II. THE STATISTICAL EVALUATION OF TRAINING

The National Training Center has been established to train the heavy battalion task force in such a way that it cannot be done at the unit home station. To achieve this end it quite rightly relies upon the train-evaluate-train model of learning as previously discussed. Although MILES and other training devices undoubtedly provide a quantum improvement in the quality of training at the NTC, such devices are being packaged for use at the home station. The key then, to the uniqueness of the NTC lies in the quality of its evaluation process. This, in effect, separates it from the usual ARTEP style training conducted at the unit home station. Given the essential elements of proper tactical training evaluation as formulated by Scott, Meliza, Banks and others at ARI, can the NTC provide information about the tactical performance of units that will bear statistical scrutiny?

Indeed, the answer to this question must be yes, for the essential methodology used by ARI to evaluate field training exists at the NTC. A hit/kill system to accurately reflect the casualties of war is utilized while a dedicated OPFOR element provides realistic opposition to the friendly unit.

A cadre of trained, professional controller/evaluators provide the necessary subjective input to the evaluation system, and are augmented by position location systems that are more accurate than the human senses. In effect, the amount of information being extracted from a unit training session is sufficiently reliable at this point to provide the basis for a proper training (learning effects) evaluation.

A completed top-down analysis of the large unit combat process also serves to provide the necessary lattice of subtasks to conduct meaningful evaluations. This "measurement of evaluation" methodology is discussed in detail in the Appendices. However, the existence of reliable data does not insure its efficient use; a methodology must exist in order to utilize this data effectively in after-action reviews (AAR) and for later provision in a take-home package to assist unit training. The current system of NTC feedback as designed has some weaknesses.

A. PROVIDING THE BATTLEFIELD: AN OVERVIEW OF THE NTC

The National Training Center concept was approved in 1977 to fulfill a pressing need within the Army. Due to rapid advances in weaponry the Army in the late 1970's found

itself unable to fully exercise units to their maximum capabilities. Realistic large-scale live fire exercises became impossible due to the range of new weapons, while the use of full electronic warfare (EW) or chemical/biological/radiological (CBR) capabilities posed dangers or unacceptable inconveniences to the surrounding civilian areas. In addition, the creation of standing OPFOR units, the use of professional training evaluators and the integration of engagement simulation technologies into large scale maneuvers proved to be prohibitively expensive if these systems were to be deployed at individual unit home stations. Thus, in order to provide these enhancements to unit training, the concept of having a National Training Center was born; all of these training requirements could be incorporated into a single station where units could be rotated to in order to train properly (see Figure 7).

1. <u>The Training Environment</u>

The objective of the NTC is to:

> "Provide a facility where heavy battalion task forces, controlling brigade headquarters, and supporting units can undergo essential combat arms training that cannot be accomplished at home stations due to physical limitations and prohibitive cost of providing a realistic training environment." [Ref. 13:pg. 1]

As a secondary objective, data from the NTC will be used to improve current Army procedures and assist in new combat

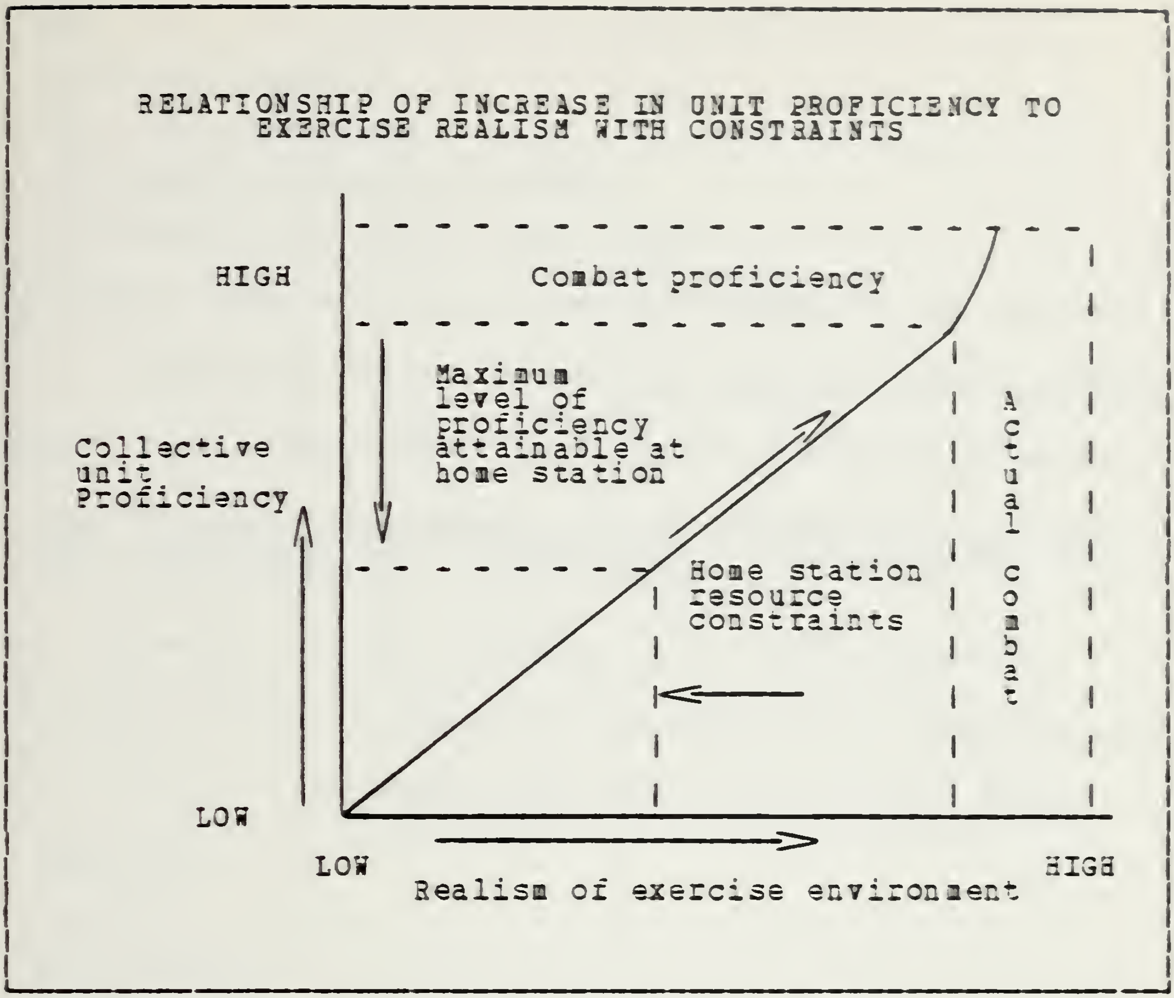


Figure 7: The Increase in Proficiency vs Exercise Realism

developments [Ref. 13:pg. 1]. Training at the NTC is based on the "Train-Evaluate-Train cycle", with emphasis on the feedback to units to provide maximum learning and to assist in later home-station follow-up training. The training emphasis is placed on heavy battalions and developing their ability to "shoot, move, and communicate." To this end the elements of the training environment are:

OPFOR

EW

Close Air Support

Live Fire Exercises

Engagement Simulation (MILES)

Instrumentation [Ref. 14:pg. I-3]

Although instrumentation is heavily relied on for analysis of the exercises and for providing feedback to the training units, the accomplishment of training itself is paramount;

"The training environment will be paramount at the NTC. Data collection will be secondary to accomplishing unit training objectives." [Ref. 13:pg. 2]

The evaluation of unit training at the NTC will be in accordance with the appropriate ARTEP with concentration on five levels of activity within the battalion task force:

Execution

Control

Coordination

Support

Planning [Ref. 14:pg. I-5]

The structure of activity for training units at the NTC is, however, unique. The missions to be performed at the NTC are drawn from Army doctrine [Refs. 15,16,17] and a complete list is provided at Appendix A [Ref. 18:pp. 1-12,1-13]. These missions (of which some are indicated as critical and will be performed by all units) are then assembled into a tailored package of scenarios which can be varied in

intensity to match the needs of the training battalion. Thus, although scenarios and intensities are pre-written at the NTC, by selecting from a "menu" of them a battalion commander and his staff can tailor their training at the NTC to directly meet their own unique needs.

The evaluation of units training at the NTC will be accomplished by the integration of several methods. Basically, they can be classed as monitering methods using the assistance of the instrumented battlefield provided at the NTC, and the traditional observer methods relying on the judgement of observer/controller personnel accompanying the unit. These two methods complement each other to a high degree and interface with each other interactively. The system serves to bridge the traditional shortfall between training constraints and evaluation necessities and provides data which can be utilized in the statistical evaluation of tactical performance.

2. Subtasking: The Top-Down Analysis

The top-down analysis technique was prescribed in 1979 as the methodology to be used in determining the Measures of Effectiveness (MOE), Measures of Performance (MOP), and Essential Elements of Analysis (EEA) that would comprise the NTC's evaluation program [Ref 14:pg III-5].

The method, generally described, is:

> "Top-Down Analysis. The technique is to assemble appropriate data sources to determine in the broadest context the missions of each system (maneuver, indirect fire, engineer). The system's missions are then decomposed into progressively lower subdivisions until quantitative measures can be applied directly." [Ref 14:pg. III-5]

This analysis, however, specifies no standards. The intent of it is to produce a listing of quantitative measures which can be used to conduct unit evaluations. This listing will later form the basis for the actual evaluation of training standards for large units.

> "The major departure from the model used in ARTEP development is that the MOE/MOP model identifies measureable elements without imposing a standard....the measures are quantitative not qualitative. The qualitative assessment will be applied after the development of quantitative measures." [Ref 14:pg. III-5]

The hierarchy for this procedure is shown at Figure 8.

The definitions for these items contained within the pyramid are at Appendix B. The procedure here is the same one as used by Scott, Meliza, and Banks to produce the statistically reliable REALTRAIN results [Refs. 8,9,10,11].

The task of performing the top-down analysis of MOEs for the NTC fell to the Combined Arms Center (CAC) at Ft. Leavenworth in 1979. Analysis began there under the auspecies of the Combined Arms Training Development Activity (CATRADA), Unit Training Directorate (UTD).

The final EEA selected for use in further subtasking were:

Maneuver
Intelligence
Air Defense Artillery
Mobility-Counter Mobility
Combat Service Support
Fire Support
Nuclear-Biological-Chemical
Command, Communications, and Control [Ref 16:pg. 3-11]

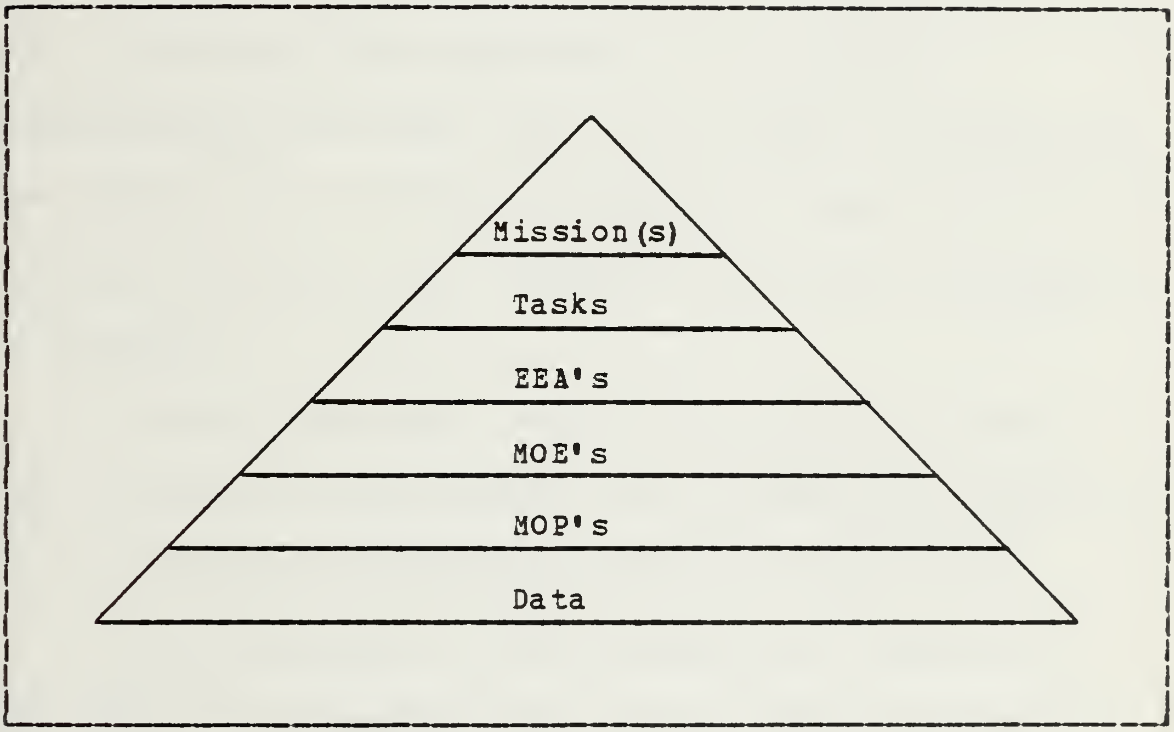


Figure 8: The Top-Down Analysis Hierarchy

For a discussion of the analysis process which resulted in these EEA, see Appendix C. These eight EEA accurately reflect the combat process of the heavy battalion Task Force (TF).

With the identification of the eight EEA, the Unit Training Directorate at CATRADA turned to the task of selecting those MOE/MOP which would support the evaluation of unit tactical performance under the EEA. Within the TRADOC community a MOE is defined as:

> "A measure of effectiveness is a quantitative indicator of the ability of a human, human/materiel, or materiel system to accomplish the task for which it was designed. For a military force, it is a measure of the ability of the force to accomplish its combat mission." [Ref 19:pg. 25]

The ideal MOE should measure directly the degree toward which a particular EEA contributes to the mission accomplishment [Ref 20;pg. 29]. As the MOE is a measure of some quantity, it should be considered in light of the four measurement scales. For a discussion of these four scales see Appendix D.

CATRADA developed a unique methodology to assist them in completing this portion of the subtasking analysis. This methodology is described in detail in Appendix E.

The culmination of this process was a series of EEAs, MOEs, and MOPs which for the first time accurately reflected the heavy battalion combat process. This subtasking analysis was then used to define the parameters of the planned NTC instrumentation system. As CATRADA now had a reasonably clear picture of the tasks to be undertaken by battalions, the instrumented battlefield--designed to

measure the accomplishment of these tasks--could now be created.

3. The Instrumented Battlefield

The data collection process at the NTC will be extensive and complex. Data will be gathered from live-fire as well as force-on-force engagement simulations. Inputs to the data collection center will be made in several forms: video recording of events; radio messages and data supplied by field controllers; monitering and recording of radio nets; and information received via the instrumented environment. The basic instrumentation concept is illustrated in Figure 9.

The Core Instrumentation Subsystem (CIS) is the computer recording center. The CIS receives all data inputs and serves as the operating area for the Exercise Management and Control (EMC) teams and Training Analysis and Feedback (TAF) teams. Antennas selectively located throughout the maneuver area provide the necessary links between the field instrumented exercise areas and the CIS. The live-fire exercise area is also instrumented to record battle data.

NTC phase I instrumentation will consist of the components shown in Figure 10. The CIS is the central computer facility that provides all real-time data

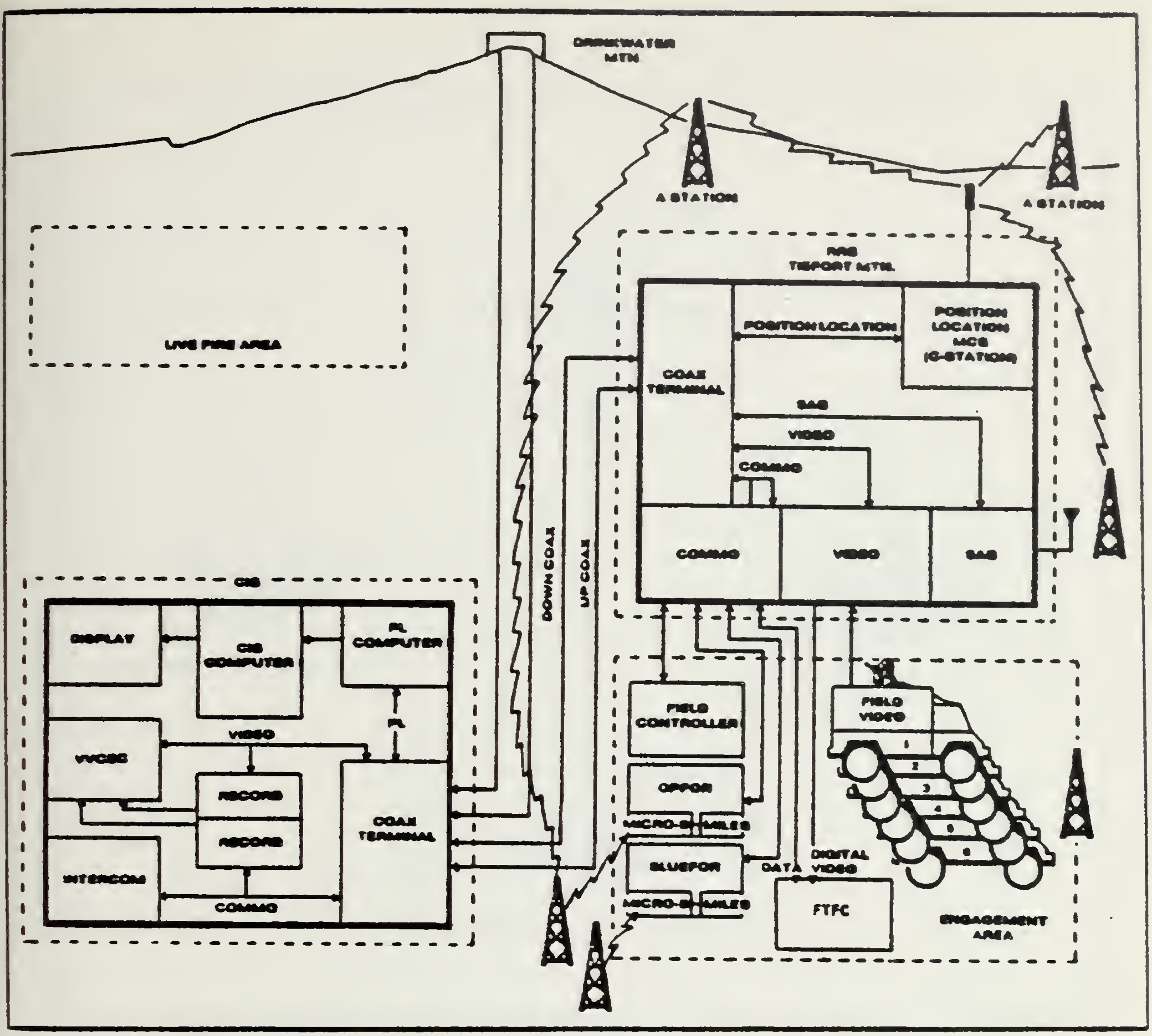


Figure 9: Pictorial Diagram of NTC Instrumentation System

processing and interactive displays necessary to moniter, command, control and evaluate the training in all the NTC exercise areas. The Range Data Measurement Subsystem (RDMS) provides real-time position location and engagement event data for all instrumented players. The Range Monitoring and Control Subsystem (RMCS) includes automated and human

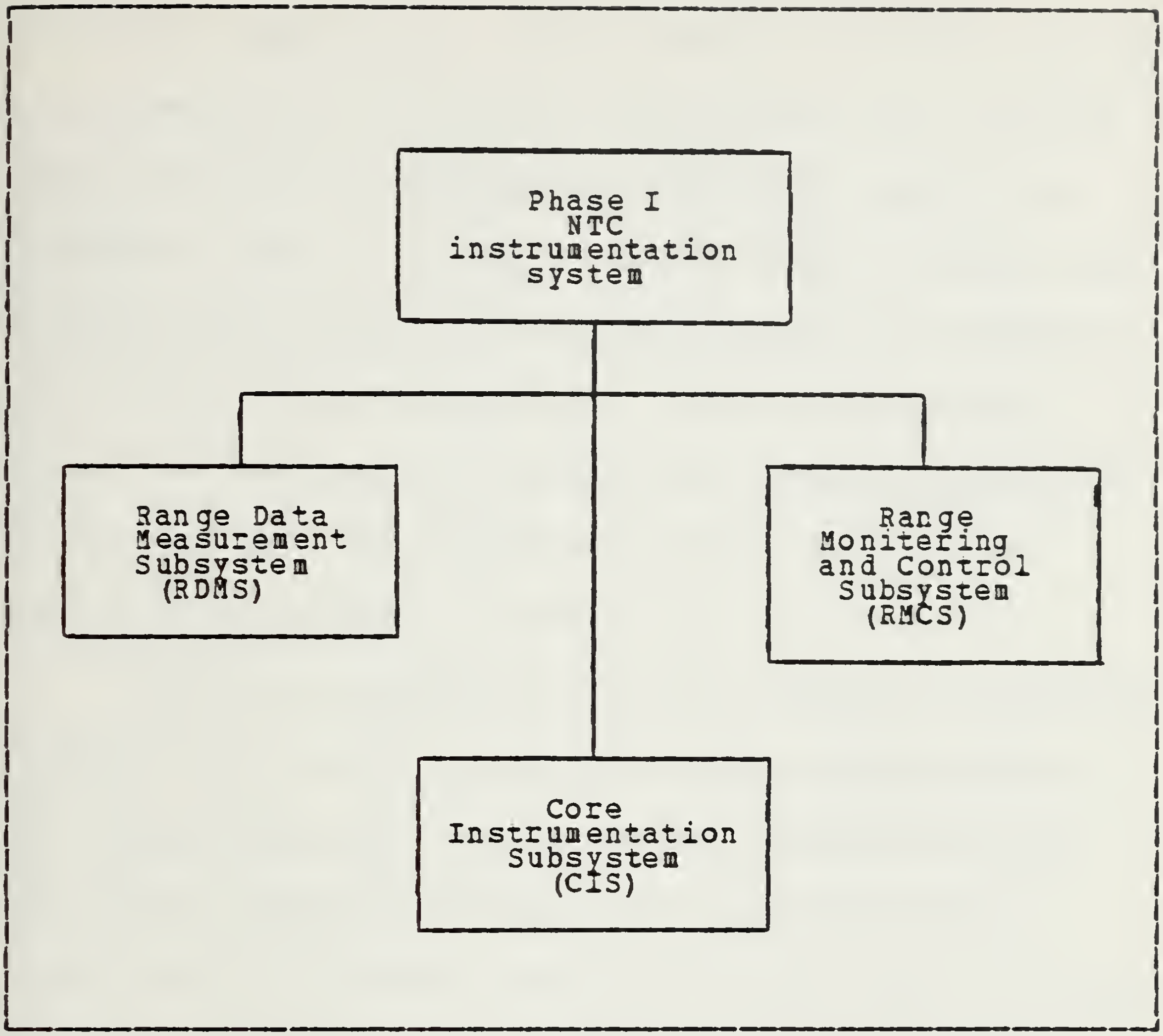


Figure 10: NTC Phase I Instrumentation System Architecture

sensors (field controllers) with communication links to the CIS; this provides the means of monitering and controlling all activities in the NTC training area. A Digital Interface Component (DIC) provides the input/output link for all digital data communications between the CIS and the RDMS and RMCS. (A more thorough discussion of the component functions is provided in Reference 18, pp. 1-14 to 1-21.)

Each subsystem has its own capabilities and functional requirements. Data gathered by the RDMS and RMCS is transmitted to the DIC. The DIC performs all electronic signal and data transformations required to provide the transmitted data to the CIS in useable form. The CIS then processes and displays the data as necessary for analysis, evaluation and decision making. Personnel in the CIS can control the exercises by transmitting messages through the DIC to the RMCS. An illustration of the data flow and control links is given in Figure 11.

Initial instrumentation at NTC consists of equipment already tested and in use by various Army experimentation and testing agencies. This equipment includes the MILES devices for casualty assessment and instrumentation to record position location, event occurrences (such as firings and radio transmissions) for 125 player items. Larger weapons such as tanks and TOWs will be fully instrumented and also interact with MILES. Future expansion will instrument 500 players, record events in more detail and will also include activities at fire direction centers, command posts and tactical air support control points as well as individual players.

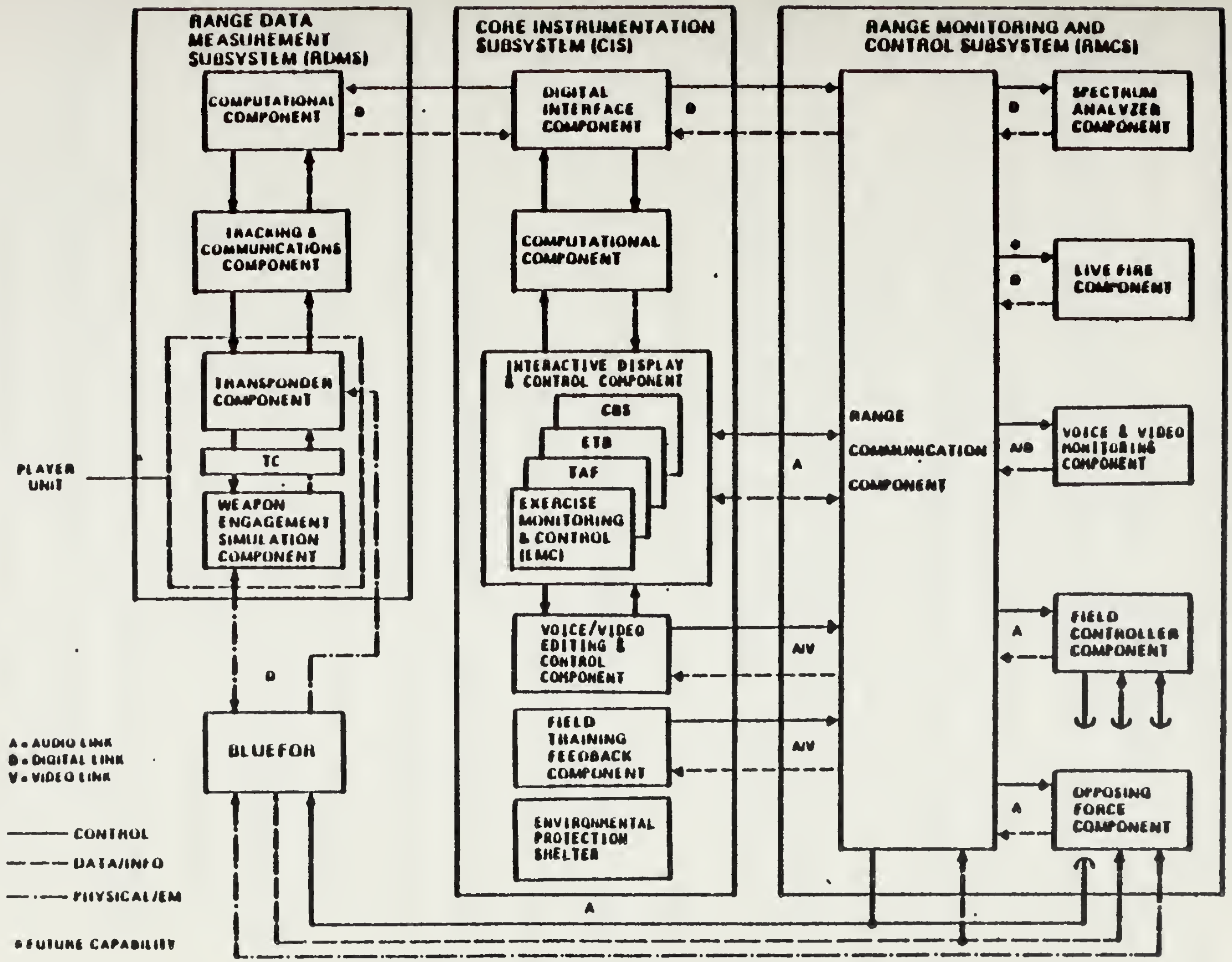


Figure 11: NTC Instrumentation System Interfaces

There are multiple capabilities contained in the Phase I instrumentation package. Various devices will input quantifiable data to the CIS. An example of this data includes time of weapon firing (each weapon recorded individually); whether a firing was a hit or miss; if a hit is generated, the individual hit is identifed and the degree of damage recorded; the positions of both players; and much more. A disscussion of the actual quantities being measured by the NTC instrumentation is provided at Appendix F.

The CIS software is programmed to tabulate and compute such relevant information as: number of rounds fired by each weapon system, weapon type, unit, etc.; number of enemy kills, by each weapon and type; range of engagements; and much other useful information. This data is then manipulated into various formats to determine the statistics used for unit evaluation: movement rate; weapons proficiency based on the number of rounds fired to kill the number of enemy targets; and so forth. A complete delineation of all statistical compilations currently planned for the NTC is provided at Appendix G.

The instrumentation has two other important capabilities worth noting. In excess of 20 different radio nets will be completely recorded during exercise periods

which will provide a means of checking key events, and there will be 6 video teams at various locations around the battlefield to film the actual exercises. The entire operating system is extremely complex and will undergo improvement as experience is gained at the NTC. As with any complex system, however, there are limitations to the quantification capabilities of the NTC. These limitations are discussed fully at Appendix H.

All data will be recorded in the CIS for evaluation, and in such detail that each training element's operations can be examined as a separate entity down to the individual firing platforms. To assist in overall unit assessment the data can be compiled and aggregated to produce battalion level statistics for whatever period is desired. The display formats and data manipulation capabilities are extremely flexible and allow a user to alter the format of the desired statistics (For a more detailed discussion of equipment capabilities and interface devices see Reference 18).

B. THE CONCEPT OF STATISTICAL EVALUATION

In discarding the old ATP/ATT system the Army effectively did away with unit "grading". This certainly was a wise decision, for the grading system of the ATT was

abused badly. Commanders often "stacked" the deck by managing crews and elements to avoid testing poor performers in order to obtain good grades. Other commanders unfairly used the grade results to deal out "punishment" in one form or another to poor performers. The ARTEP philosophy of evaluation (and subsequent re-training and improvement) without grading is sound and must be maintained. Any such proposed methodology must avoid the appearance of a "report card" or a comparison between specific units.

Yet clearly, there must be a form of differentiating performance. "Mission accomplishment is too vague a criteria, for certainly a unit which "takes the hill" with 85% casualties has accomplished its mission just as surely as one which "takes the hill" with 20% casualties. The Army, however, must discriminate between the trained and the untrained; two such units obviously differ (all other factors being equal) in their acquired combat skills. This problem refers back to the major weakness of the ARTEP itself: the lack of a quantifiable standard for evaluating the tactical performance of large units. In order to quantify such an evaluation, some sort of measureable standard (not a grade) becomes necessary--be it percent casualties, loss-exchange ratios, targets killed per rounds

expended, or some other such measure of effectiveness. Any program which serves to clarify these parameters will ultimately help to refine the ARTEP itself into a more useful document.

1. Evaluation Without Grading

Fortunately, there exists such a methodology for examining performance. In academic circles during the early 1960's methods were devised to test students on their intrinsic academic skills, without tying such an evaluation to a series of grades. Most students are familiar with these results; the Scholastic Aptitude Tests (SAT), various College Entrance Examination Board (CEEB) tests, the Graduate Record Exam (GRE), and in the public schools the Iowa Test of Basic Skills (ITBS), all utilize essentially the same concept in evaluation, and it is a simple one. Basically, the idea is that a student's performance--based on the numerical outcome of a given test--can be compared to a large, anonymous population (in effect, a normal population), and using the method of z-scores, t-scores, stanines, or any one of several statistical methodologies, a percentile rating can be calculated for the individual's performance in that skill area. Thus, instead of being told he received a "B" in mathematics, a student is told he falls

in the 80th percentile for his age group--a much more meaningful assessment, for it tells the individual his relative skill as it relates to his peer group.[5]

This method has long been employed in the US military to classify soldiers for service potential, which is a situation where grading has little or no usefulness. In a similar fashion it could be used to relate tactical proficiency of a unit to its commander and staff without resorting to any grades or unit comparisons. Certainly an evaluation such as "you took the hill and suffered 20% casualties; this performance falls in the 84th percentile of all units who have accomplished this training objective" is more informative to a commander than a brusque "you took the hill and lost 20% of your force." One statement relates a degree of accomplishment, while the other provides no hint regarding whether or not 20% casualties was historically excessive for this operation.[6]

----

[5]It must be stressed here that this is an anonymous peer group. No comparison of "Jimmy vs Johnny" is implied, or, in fact, is even possible under this methodology.

[6]Indeed, the NTC development plan [Ref. 14] requires the determination of quantified (historical) standards for each scenario, and states that:

> "A secondary objective will be to define baseline (norm based) standards (for scenarios) by trend analysis of the data base." [Ref. 14:pg. III-10]

This has not been accomplished as yet.

Other benefits accrue automatically from such an evaluation. Using a data base of these evaluations, it would be possible to analyze them for inherent doctrinal weaknesses; for example, if a particular operation yields excessive casualties repeatedly, it may be determined that this particular combat skill needs greater emphasis throughout the Army, or that our doctrinal approach to the problem is flawed. Further side effects of such an evaluation system would be increased insight into the combat process, and ultimately an enhanced ARTEP--one containing more specific evaluation criteria which could ultimately save lives in combat.

2. The Training Readiness Profile

The Training Readiness Profile (TRP) about to be proposed here would provide a vehicle whereby a unit evaluation could be constructed in a non-competitive manner without resorting to grades. Indeed, any other sort of historical comparison is expressly forbidden:

> "Unit evaluations will be for the sole purpose of facilitating remedial training of the unit. There will be no unit comparative scores or scoreboards."[Ref. 13:pg. 2]

The TRP would be simple, small, easy to read, and hopefully more informative to a battalion commander and his staff than a series of bare statistics presented in a disjoint fashion

at an after action review (AAR). The TRP could begin with a chart depicting overall mission accomplishment in a percentile fashion calculated for each major combat mission (Appendix A) conducted while training at the NTC. This item would provide a basic overview of the unit's tactical performance across the board. Following mission accomplishment, pages depicting the unit's ability to shoot, move, and communicate in each training mission would be displayed. Finally, feedback from the observer/controller evaluations could be portrayed followed by pertinent comments from the EMC/TAF personnel, the Commander of the NTC, and other individuals so tasked to provide a unit evaluation. These comments would serve to counterbalance and augment the machine generated statistics and introduce a human (albeit more subjective) evaluation of unit performance.

Such a document fills a need in the currently planned system of evaluation and feedback at the NTC. The after action review (AAR) methodology now planned for the NTC is somewhat incomplete; although it is efficient in reflecting what happened during a training exercise, and provides some insight into why such things happened, it does not convey degree of performance to the commander--how well

he did is left for the commander to surmise on his own.[7] Any allocation of training resources at the home station following an NTC training cycle is then done based on his perception of what was acceptable or unacceptable and any linkage between certain levels of performance and his overall unit mission accomplishment is left to the commander's own insight. This is perhaps not the most efficient use of the data generated at the NTC. In all likelihood, a battalion commander and his staff will only go through an NTC training cycle once together (planned rotation cycle is every 18 months for each battalion at the NTC). In order to capitalize on the experience and improve the conduct of home station training, some degree of relative importance must be attached to training deficiencies highlighted by the NTC exercise. The proposed TRP does this in a concise and economical way, for a commander whose unit performed in the 20th percentile for communications procedures in the hasty attack and suffered correspondingly high casualties while remaining in the 80th percentile for other MOEs can draw a conclusion much more

----

[7]In conversation with EMC/TAF personnel at the NTC, there was no indication that the initial standards had yet been generated for scenarios as required by the NTC Development Plan [Ref. 14:pp. III-7 to III-11]. This perception could be erroneous, but even so, the TRP would provide an automatic methodology to implement these instructions.

accurately than a commander who is only given a numerical summary of communication transmissions without knowing the impact or whether the number of tranmissions was excessive. Accordingly, the first commander who has availible for his use a percentile standing knows his unit is below the mean (average) in this area and can focus his efforts to remedy this shortcoming. The second commander only knows he suffered a large number of casualties and likewise knows the number of radio transmissions made, but he does not have the obvious link between the two facts since no indicator of average performance is provided.

Thus, the TRP possesses several advantages (which will be examined in detail later). It reflects true unit readiness, that is, it applies a relative standing within the peer group to specific training performance. Using this method, if one desired, true aggregated readiness statistics could be generated for units within the Army that would reflect a common assessment. The proposed format is designed to be easy to understand and would not require excessive statistical reports. Virtually all officers should be readily familiar with normal score percentiles, as all college entrance exam results are in this format. The TRP would also be a true resource allocation tool--it would

demonstrate the need for remedial training clearly and provide an easy method for linking cause (training deficiency) and effect (casualties).

Finally, given a large enough population for examination, the TRP could be used to set defacto standards for the ARTEP. After the data base has become sufficiently large to permit the assumption of normality, statistical tools could be applied to such information to yield quantifiable standards in unit training. This could improve the ARTEP and yield an even greater benefit to the Army.

In this document the methodology used to establish the current subtasks in use at the NTC will be examined, and changes proposed to permit the generation of a TRP. The theory behind constructing the TRP will be discussed and followed with a sample TRP for examination. Lastly, a discussion of the possible growth and implementation of this system will be followed by a series of recommendations and conclusions regarding the TRP system. In total then, this document shall demonstrate an improved methodology for the evaluation of unit tactical proficiency at the National Training Center.

III. THE MEASUREMENT OF TRAINING ACHIEVEMENT

The key, as stated previously, to the uniqueness of the National Training Center experience lies in the quality of its evaluation process. First and foremost the NTC exists to provide training to the heavy battalion task force. However, in using the "Train-Evaluate-Train" cycle the advantage of training at the NTC can be squandered if the feedback process--the evaluation stage-- is inadequate to the task. In order to capitalize upon the first-rate training opportunity provided, the most effective use must be made of the statistically reliable data being taken from the instrumented training environment.

The purpose of the TRP is to provide an objective normative evaluation of unit training performance. This method is the best solution currently available to the problem of evaluating training standards for large units. The TRP will be generated for each exercise segment, which is generally comprised of one mission (see Appendix A). Any evaluation to be conducted must be based on measures concerning the activities that transpired during the segment. Since these measures are being compiled in order to produce an objective evaluation, the measures themselves

must be objective. The goal, therefore, is to develop a set of objective measures that reflect overall training performance. As defined by TRADOC, an MOE is a quantitative indicator of the ability of a military force to accomplish its combat mission [Ref. 19:pg. 25]. The issue now at hand is to determine and define a set of MOEs that meet these requirements.

Any set of MOEs that satisfy the above restrictions could be used to produce an objective training evaluation. There is no unique set that is optimal, but certain measures are generally considered to be "better" than others in providing a basis for assessment. The collective set of MOEs should reflect overall unit effectiveness and not just a series of disjoint "snap-shot" looks at specific areas of performance (i.e., the set must cover the full spectrum of performance to be evaluated). The MOEs should be defined in such a way that different beginning force strengths and capabilities are accounted for in the resulting number, i.e. the measure should be "normalized". Percentages provide this normalizing factor, but interpretation or comparison of percentages can be misleading if the actual inputs to the percentage computation are not available. In most instances

a larger number is associated with being better so the measures should be defined with this in mind.

The NTC collects an array of objective data from training exercise activities. The collected data and resulting statistics have formed the basis for developing a set of MOEs that can be used to provide an objective evaluation of the unit's training performance. The remainder of this section will present a set of MOEs that provide a broad coverage of the unit activities. MOEs fit into the current stucture of analysis categories at the NTC: mission accomplishment; shoot; move; and communicate. Collectively, the MOEs yield an overall assessment of the unit's performance. Where percentages are used to normalize portions of the data, the actual numbers used in the computation will be given. In all cases, except where specifically noted, a larger number represents a higher level of performance.

A. MISSION ACCOMPLISHMENT

A tactical mission is a tasking to perform a specific function, i.e, attack, defend, etc. (see Appendix A). Other instructions in the mission usually include such details as when, where, or why the task is to be accomplished. There is no single set of guidelines or standards that can be used

to judge whether or not an assigned mission was successfully performed. In some cases this leads to a somewhat ambiguous and subjective process in determining success or failure for a mission. However, there are general areas of agreement when assessing mission performance. These areas are as follows: the friendly force must have enough survivors to continue operations; the enemy force must be destroyed or weakened to hamper further operations; and, if a time is specified for an operation, it must be met. All of the following proposed MOEs were developed so as to provide flexibility in interpretation in these areas without imposing a predetermined standard of acceptable performance. Individually, the following MOEs provide an assessment of unit performance in a limited view, but collectively they represent the measures necessary to evaluate overall unit performance.

1. Percent OPFOR Vehicles Killed (POVK)

This is the percentage of OPFOR combat vehicles that are killed by friendly weapons .

$$POVK = \frac{\text{\# of OPFOR combat vehicles killed}}{\text{\# of OPFOR combat vehicles in initial force}} \times 100$$

This is a direct measure of the mobility and firepower loss suffered by the enemy. It addresses the effectiveness of

the friendly force in directing fire at and/or destroying enemy vehicles. Since a percentage is used here the measure reflects loss relative to a particular starting strength.

2. Percent OPFOR Personnel Killed (POPK)

This is the percentage of OPFOR personnel that become casualties.

$$POPK = \frac{\text{\# of OPFOR personnel killed}}{\text{\# of OPFOR personnel in initial force}} \times 100$$

This is a direct measure of friendly killing power effectiveness and reflects the loss strength suffered by the enemy. "(This) measure is used to evaluate total force effectiveness when destruction of the opposing force is the primary mission of both sides." [Ref. 21:pg. 4-43]

3. Percent OPFOR Loss Value (POLV)

This is the percentage of the total value of the OPFOR that is killed by the friendly force during the exercise segment.

$$POLV = \frac{\sum (\text{\# of type "i" OPFOR targets killed}) \times (\text{value of target type "i"})}{\text{value of initial OPFOR unit}} \times 100$$

The value of each OPFOR type is provided in the weighted effectiveness index (WEI/WUV) table stored in the CIS (see Figure 22, Appendix G). This measure combines the heterogeneous force structure of the OPFOR in a fashion that

reflects the total fighting capability of the OPFOR that was destroyed. The enemy usually does everything possible to thwart the accomplishment of the friendly force mission. This is a weighted measure which depicts the amount of enemy combat power destroyed by the friendly force.

4. Percent Friendly Vehicles Survived (PFVS)

The percentage of friendly combat vehicles that are not killed by enemy fire.

$$PFVS = \frac{\text{\# of friendly combat vehicles surviving battle}}{\text{\# of friendly combat vehicles in initial force}} \times 100$$

This MOE is widely used to provide a direct measure of the survivability of the friendly force. It is an indicator of how well the friendly force used available terrain to mask movement and protect vehicle positions. A unit's mobility is highly important in the battle area. This measure reflects how well the friendly force was able to utilize its mobility and conserve its combat power.

5. Percent Friendly Personnel Survived (PFPS)

This is the percentage of friendly personnel who survive the battle.

$$PFPS = \frac{\text{\# of friendly personnel alive at end of battle}}{\text{\# of friendly personnel in initial force}} \times 100$$

This is an important measure which indicates whether or not a force is capable of continuing operations. Successful mission accomplishment is highly questionable if the friendly force is reduced to a negligible strength during a battle. This MOE reflects how well the friendly force avoided enemy fire and is an indicator of the unit's ability to properly execute sound tactical plans and conserve the lives of its soldiers.

6. Percent Friendly Survival Value (PFSV)

This is the percentage of the total value of the friendly force that survived the enemy fire.

$$PFSV = \frac{\sum (\text{\# of type "i" friendly forces alive}) \times (\text{value of friendly type "i"})}{\text{value of initial friendly force}} \times 100$$

The value of each friendly type is provided in the WEI/WUV (see Figure 22, Appendix G) table stored in the CIS. This MOE combines the heterogeneous structure of the friendly force to reflect its survivability. Besides inflicting enemy casualties and capturing objectives, the friendly force must protect its personnel and equipment so they can continue to fight. This weighted measure reflects how well the unit conserved combat power.

7. <u>Relative Loss Exchange Ratio (RLER)</u>

This is the proportion of OPFOR losses divided by the proportion of friendly losses.

$$RLER = \frac{(\ POLV\ )}{100-(\ PFSV\ )}$$

The measure will be a pure number that compares the relative value of the losses suffered by both sides. The relative loss exchange ratio is a measure of friendly effectiveness taking two major factors into consideration. The numerator reflects the destructive capability of friendly weapons. The denominator is a reflection of the ability of the friendly force to survive. A combination of these factors in a ratio approaches an overall indication of friendly force combat effectiveness. [Ref. 21:pg. 4-22]

8. <u>Time to Accomplish Mission (TAM)</u>

This measures the total elapsed time from beginning to completion of a mission.

TAM = (mission end time) - (mission begin time)

As a stand alone measure this MOE would not be very useful. However, when a time is specified for a particular mission this measure becomes vitally important. Even if a time is not specifically stated this MOE can reveal deficiencies in planning, movement and overall mission execution. For this

measure, the stated mission will dictate whether a large or small number is better. In an attack the friendly force wants to capture an objective quickly before the enemy can block the advance. For a delay or defense the friendly force aims at holding back the enemy as much as possible so a longer time is better.

B. SHOOT

The friendly forces must effectively and efficiently employ their weapons to gain the maximum benefit of their firepower. Direct and indirect fire must be organized and placed in such a manner as to suppress the enemy capabilities of fire and detection. Weapons must be used to conceal and protect friendly movement as well as to destroy the enemy force. To accomplish its mission, the friendly force must "destroy enough of the enemy to convince him to break off his attack, to give up a defensive area, or to move from an area vital to friendly forces." [Ref. 16:pg. ii]

> "The tank with its cross-country mobility, its armor protection, and its formidable firepower, has been and is likely to remain the most important weapon in the battalion task force. The accuracy of tank guns gives them a high probability of a first round hit and the lethality is such that if the target is hit it will be killed." [Ref. 16:pg. 1-2]

The tank and other modern weapons require trained operators to realize the full potential of this firepower. If aiming

and firing are not done correctly large amounts of ammunition will be expended with only a minimal number of enemy casualties resulting. If gunners incorrectly identify targets or use poor target selection procedures the friendly force could sustain more casualties than necessary while the enemy continues to fight. An important point to remember is that enemy and friendly weapons possess similar capabilities so a major contributor to mission success or failure will be how well each side uses its firepower potential.

To assess the use of firepower potential the performance of the weapons systems must be examined. Each weapon system has its own individual impact on battle outcome. Along with this impact is a synergistic effect gained from the mutual support of other weapons. The selected MOEs for mission accomplishment measure this combined force effect and reflect the overall force effectiveness. This category of MOEs is concerned with assessing the performance of selected weapons systems: tank main gun, TOW, and Dragon. The performance of each system will be displayed separately in the TRP, but will be measured using the same MOEs.

Key areas in assessing weapon performance are the accuracy and lethality of the weapon, its contribution toward mission accomplishment, and the amount of ammunition

expended by the system. The latter factor impacts heavily on the required logistic support. Proper tactical employment of all sytems must be considered, but note that this overlaps with the categories of mission accomplishment and movement. Shoot MOEs are defined to address these issues as an input to performance assessment.

1. Number of Rounds Fired (NRF)

The total number of rounds fired by all weapons of this type.

$$NRF = \sum \text{(all rounds fired by this weapon type)}$$

This measure reflects the fire capability of the particular weapon system. It is also used to assess ammo conservation and accuracy. This is not a stand alone measure and must be considered along with the enemy casualties inflicted. This number does, however, provide an indicator for logistical requirements and when combined with the number of casualties reflects weapons proficiency. A small number in this MOE is good, provided the mission is successfully accomplished.

2. Casualties Per Round (CPR)

A ratio of the number of enemy tank and antitank weapons killed by this weapon type to the total number of rounds fired by this weapon.

$$CPR = \frac{\text{\# of enemy tank and antitank weapons killed by this weapon type}}{\text{\# of rounds fired by this weapon type}}$$

"This measure addresses kill productivity of a weapons system directly." [Ref. 21:pg. 4-142] It considers both the accuracy of a weapon and its lethality.

3. Percent Rounds Hit Target (PRHT)

The percentage of the total number of rounds fired that hit a tank or antitank weapon system.

$$PRHT = \frac{\text{total \# of hits on a tank or antitank weapon system by this weapon type}}{\text{\# of rounds fired by this weapon type}} \times 100$$

"This is a direct measure of accuracy of fire." [Ref. 21:pg. 4-159] There are some items that should be considered when using this MOE for an assessment. When accuracy of fire appears to be poor the range between the firer and the target must be examined. A problem of attempting to use a weapon past its maximum effective range is somewhat different from missing targets that are within range. This MOE does not distinguish between these two points. Also, the NTC instrumentation currently uses hits and kills as two distinct and mutually exclusive groups of weapon effect. To derive the numerator of this MOE requires that these two groups be summed.

4. Percent Rounds "Near Miss" (PRNM)

The percentage of the total number of rounds fired that were close enough to the target to impart a suppressive effect (and be reflected by the MILES instrumentation), but did not actually hit the target.

$$PRNM = \frac{\text{\# rounds counted as near miss against tank and antitank weapons by this system}}{\text{\# rounds fired at tank/antitank by this weapon system}} \times 100$$

> "Near misses are thought of as a suppressive effect. They might also be thought of as a secondary guage of accuracy of fire in the sense that a system with a high percentage of near misses is more accurate than one with more outright misses, when percent hits is equal." [Ref. 21:pg. 4-160]

This measure is not used alone, but should be used with the other MOEs as an input to total firepower effectiveness.

5. Weapons Fractional Kill Effectiveness (WFKE)

The percentage of the total enemy casualties that was inflicted by this particular weapon system.

$$WFKE = \frac{\text{total value of enemy casualties inflicted by this weapon type}}{\text{total value of enemy casualties inflicted}} \times 100$$

The value of each enemy player is available in the WEI/WUV table (Appendix G) stored in the CIS. This measure reflects the relative effectiveness of the major weapon systems in the force. It portrays how much each weapon system contributed to the total casualties suffered by the enemy.

When combined with the other MOEs it allows for assessment of tactical weapons employment and firepower effectiveness of each weapon system.

C. MOVE

On a modern battlefield of highly lethal weapons the friendly force must be proficient in maneuver in order to survive. The enemy can detect movement that is not masked by terrain or deception. A friendly force must conduct a thorough map study and ground reconnaissance when possible before selecting routes. The OPFOR is extremely mobile, and is capable of moving quickly to block friendly advances or to attack through gaps that might occur between friendly forces.

To be successful, the friendly force must be capable of moving quickly to engage a retreating enemy before he can reorganize. If a gap exists, units must maneuver promptly to prevent an enemy penetration. The keys to success include using the terrain properly and the capability to move forces quickly to decisive places in the battle area.

Measures in this area of performance should reflect the ability of units to meet the above requirements. A leader who uses sound tactical principles will select routes that provide protection for his forces and allow sufficient

maneuver area to use his mobility potential. A well trained unit will generally move more rapidly than a poorly trained unit because subordinate elements will require less direct control and supervision. When vehicles are properly maintained they operate longer before breaking down, thus providing more mobility potential to the force.

The ability to move efficiently and effectively is of vital importance. Fast movement rates do not necessarily reflect better performance. A rapidly moving force might sustain a high level of casualties because it is moving too fast to use available terrain for protection. The point to be made is that movement is an input to overall unit performance, but must be considered in the light of other factors (such as number of casualties), to provide a measure of unit effectiveness.

1. Mean Rate of Travel (MRT)

The computation of distance travelled per designated time period.

$$MRT = \frac{\text{total distance travelled}}{\text{total elapsed time for travel}}$$

As the elapsed time increases, i.e. for an entire exercise segment, this measure is very useful. This MOE will reflect the travelling speed for the unit which averages the

changing rates that occur throughout the time period. In an attack mission this MOE is important since commanders need to know how quickly their units can be expected to move to an objective. For a defense mission this MOE is not meaningful. However, in a delay this measure could be used to account for the enemy mean rate of travel; in this case a smaller number would be better than a large number. This MOE is a direct measure of movement performance. "It is considered superior to simple amount of advance which does not take into account a possible increase in difficulty of advance as distance from enemy decreases." [Ref. 21:pg. 4-15]

2. Rate of Advance Toward Objective (RATO)

This is the rate at which a unit advances toward a designated location (objective).

$$\text{RATO} = \frac{\text{distance from start point to objective}}{\text{elapsed time to travel from start point to objective}}$$

This measure addresses the timeliness aspect of maneuverability. It is different from movement rate in that the only distance considered in the computation is that portion of movement that gets the force nearer the objective. A long, circuitous route might allow for a faster mean rate of travel than a more direct avenue of

approach, but there will be a longer total distance to move. A direct route is shorter, but may allow only a slow rate of travel. This measure combines both route selection and speed into a single number. The distance from the start point to the objective is fixed so the actual route travelled does not change the numerator, but the travel time from start to end will vary depending on the length of the route and how rapidly a force can move along the selected route. This MOE can be used to assess tactical movement plans and the maneuver capability of the force.

Since each company size element is a separate entity, usually with its own objective, this MOE will be computed for each maneuver company. This measure is meaningful for the friendly force when the mission includes movement, such as attack. In this case a larger number is better. When a friendly force has a mission to defend or delay its movement rate will be less meaningful. In this case the MOE should measure the enemy's rate of advance toward the friendly position. Again, a separate measure for each company would be appropriate. A smaller number represents better performance in this case since the friendly force is attempting to impede the enemy advance.

3. <u>Average Percent Force in Contact (APFC)</u>

This is the percent of the friendly force that is engaged with the enemy, averaged over the duration of the exercise.

$$\text{APFC} = \frac{\sum \dfrac{\text{\# of friendly platoons engaged each period}}{\text{total \# of friendly platoons}}}{\text{total \# of periods where engagement occurs}} \times 100$$

A platoon is considered engaged for an update period (every five minutes) when a weapons pairing event occurs for at least one member of the platoon. Regardless of the number of pairings each platoon will be counted as engaged only once during each period. A unit that goes through a series of periods with no enemy contact will not necessarily receive a lower number than a unit constantly engaged because only the periods where an engagement occurs are counted in the denominator.

This MOE provides a relative display of how well a commander maneuvered and positioned his force in order to direct maximum firepower at the enemy force. Where the mission is to avoid decisive contact, such as delay, a smaller number for this measure would be better provided the friendly force delayed the enemy movement as directed by his mission. This measure provides a key input for assessing

overall deployment and maneuver of forces when coupled with the MOEs for mission accomplishment.

4. Operational Readiness (OR)

This is the percentage of friendly combat vehicles that successfully participated in the exercise without a disabling maintenance failure. A disabling maintenance failure shall be defined as any failure which renders the vehicle reportable on the DA form 2406 (Materiel Readiness or Deadline Report).

$$OR = \frac{(\#\text{ vehicles begin exercise}) - (\#\text{ vehicles lost for maintenance})}{\#\text{ vehicles begin exercise}}$$

The "downtime" or time taken to repair the vehicle is not a factor here, but rather only the occurrence of such a failure in the vehicle. A unit must conduct routine and proper maintenance on its equipment if the items are expected to continue operating. Without vehicles a friendly force loses effectiveness in mobility and firepower. When a vehicle experiences maintenance failure it not only decreases the unit combat potential, but also becomes a burden for the logistics and support elements.

The number generated by this MOE is a direct assessment of the unit's vehicle maintenance capability. Vehicles lost to enemy fire are not considered to be a

maintenance failure unless the failure occurred before receiving enemy fire. No individual vehicle may be counted more than once during a segment for this MOE. This is to preclude negative OR values, which would be meaningless.

D. COMMUNICATE

The ability to communicate with other friendly elements during a battle cannot be overemphasized. Effective command and control are not possible if the leaders are unable to communicate with their units. Commanders need the ability to direct friendly maneuver and sometimes change designated plans. To facilitate quick reaction a unit must receive orders in a timely manner. When enemy contact occurs a leader is aided by indirect fire weapons if he can call an artillery or mortar unit. Coordination with adjacent and supporting elements is necessary to insure efficient employment of assets and to avoid interfering with other friendly operations. Without effective communications a battle could be total chaos.

As a point of assessment, well planned operations usually provide contingencies for possible events, thus requiring fewer changes to orders. Units that are properly trained can react to a developing situation without close direction and supervision from seniors. A commander

preplans fires to support his operation, thus requiring less time to make a call for fire when needed. When units have trained together and are more proficient in tactical operations the amount of coordination effort is reduced. Communication procedures emphasize that messenger or wire linked telephone is better than the radio whenever possible to preclude interception or jamming by the enemy. When radios are used the operators should be trained to avoid communication security (COMSEC) violations, i.e. excessive transmission times, transmitting compromising information about friendly forces, etc. Good commanders will use their available time wisely for planning and will allow sufficient time for subordinates to conduct their planning. The dissemination of orders and combat intelligence is important to insure that all friendly forces have the maximum amount of information possible to help in planning and/or executing assigned missions. These areas should be considered when assessing a unit's communication capability.

All MOEs in this category have been defined so that a smaller number represents better performance. Each MOE is aimed at providing an assessment of the unit performance for the considerations discussed above. The collective set of MOEs will reflect the unit's overall communication

capability. Any specific MOE with a high number in this area could be an indicator for identifying training deficiencies.

1. Average Transmission Duration (ATD)

The average duration of a radio transmission made by the friendly force.

$$ATD = \frac{\text{total summed duration of all friendly radio transmissions}}{\text{\# of friendly radio transmissions}}$$

This measure provides an indicator of how long each radio transmission lasted. For well trained units all communications should be brief, with the unit relying instead on standard operating procedures (SOPs). Lengthy information should be sent by some means other than radio. The longer a radio set remains keyed, the greater the chance that the enemy will intercept the signal. Communications of short duration serve to deny the enemy information about the friendly force and increase the chance of survival.

2. Average Number of Transmissions (ANT)

This is the average number of radio transmissions per time period made by the friendly force during the exercise.

$$ANT = \frac{\text{\# of friendly radio transmissions}}{\text{elapsed time of the exercise (min or hr)}}$$

As discussed above, a unit should attempt to minimize its use of the radio. An excessive number of transmissions allows the enemy to track friendly movement based on emitted radio signals. The more often a radio is used the more chances the enemy has for exactly locating the unit. The opportunity to commit COMSEC violations increases. A large number for this MOE reflects extensive supervision or direction between elements which should not be necessary for well trained units. Dividing by the elapsed time of the exercise serves to "normalize" the number for comparison with exercises of different duration.

3. Percent of Transmissions Possible RDF (RDF)

The percentage of friendly radio transmissions that exceeded 25 seconds, but were less than 55 seconds in duration.

$$RDF = \frac{\#\text{ of transmissions } >25 \text{ seconds and } <55 \text{ seconds}}{\#\text{ of friendly radio transmissions}}$$

A radio transmission period of 25 to 55 seconds is sufficient time for the enemy to "radio direction find" (RDF) a unit location [Ref. 22:pg. 57]. This means that the enemy can establish a close approximation for the friendly unit location if a radio transmits for that duration. If a unit continually transmits for this lengthy duration it will

be located and the most likely result is that enemy fire will be directed on the position. To increase the unit survivability and deny the enemy information about friendly locations this number should be low.

4. Number of Significant Transmissions (NST)

The total number of radio transmissions of duration greater than or equal to 55 seconds.

$$NST = \#\text{ of radio transmissions} \geq 55 \text{ seconds}$$

The enemy has the ability to pinpoint a friendly location if a radio transmission from that location lasts 55 seconds or more [Ref. 22:pg. 57]. To provide maximum protection for the unit the number of transmissions in this category should be minimized.

5. Percent Planning Time Forwarded (PPTF)

The "percentage of total planning time available that an echelon allows to all lower echelons." Here, the percent amount of time that the battalion task force allows to its companies.

$$PPTF = \frac{\text{time Co ordered to start execution} - \text{time TF order issued}}{\text{time Co ordered to start execution} - \text{time mission received by TF}} \times 100$$

This measure accounts for the timeliness of the command function. It considers planning time, decision time, time

to prepare an order and then disseminate it to subordinate elements. "The measure addresses effectiveness of command and control by assessing how quickly planning is completed on an order issued in relation to the time available." [Ref. 21:pg. 4-120] This MOE not only addresses the effectiveness of the command functions, but also assesses the communication procedure in the coordination phase of the planning and in the actual dissemination of the order. A rule of thumb states that each echelon should allow 50% of the available time for its subordinate units in the order preparation activities. This measure does not assess the quality of the planning or the order that was issued. [Ref. 21:pg. 4-120] The input information for this MOE must be obtained manually by the observer/controllers (OC's).

6. Mean Dissemination Time (MDT)

The time required to disseminate an order, directive, or warning to all elements at the next lower echelon of command.

$$MDT = \frac{(\text{time Bn TF order acknowledged by last Co}) - (\text{time order was approved})}{\#\text{ of orders issued}}$$

If companies are to conduct proper planning and preparation for operations they need information on a timely basis. Units that are well trained usually have SOPs for preparing

and distributing orders to subordinate elements. These methods decrease dissemination time, increase the reliability of message receipt and usually avoid radio transmission since the duration would be lengthy. This measure assesses the effectiveness of the communication link that ties the companies to the battalion headquarters. A smaller number is better for this MOE [Ref. 21:pg. 4-122]. The input information for this MOE must again be obtained from the OC. It may also be possible to obtain this information from the monitored battalion command net, if orders are habitually sent and acknowledged by radio.

IV. THE TRAINING READINESS PROFILE

Army training evaluations have never before provided statistically reliable data (see Chapter 1, section C). One of the major weaknesses of the ARTEP system has been the lack of consistent, reliable evaluations. This has promoted the vicious cycle of sister battalions evaluating each other in training, and therefore causing all such evaluations to sink to the "lowest common denominator". One unit cannot evaluate another beyond its own skill level. The NTC provides the Army with the first real opportunity to establish a consistent Army-wide training evaluation that reflects true unit performance. This opportunity will be lost if no standards, normative or absolute, are applied in the evaluation. At this point in time it is very difficult to generate absolute, fixed numbers to be used as standards for the various MOEs defined previously.[8] Without such fixed standards there is no current method available to convey the existence of a standard or level of performance to a unit. A normative evaluation, however, provides a means for solving this problem.

---

[8]Indeed, due to the synergistic effects of combat, it may be impossible to set numerical standards for some of the MOEs generated in the top-down analysis.

The purpose of the TRP is to provide an objective training evaluation to the heavy battalion task force commander in order to best facilitate remedial training in the most efficient manner. In the previous discussion of the POT cycle it was pointed out that the key to training was the establishment and evaluation of training standards. The TRP system is designed to accomplish the evaluation of training standards for large units. It does this without setting an arbitrary quantitative standard by utilizing a normative (relative performance) based evaluation. This goal is consistent with the ARTEP philosophy of training, the purpose of the NTC, and the NTC development plan as written by HQ, TRADOC. [Refs. 6,13,14]

The TRP is based upon quantified measures specifically designed (as previously shown) to eliminate human bias and error in evaluation. Its use will make substantial progress in eliminating subjectivity in performance evaluation. The format, it is felt, is concise, brief, and easy to understand. It employs as a methodology the simple concept of normal scores which is readily familiar to the Army Officer Corps. All data is presented in a format which has many intrinsic benefits, is simple to calculate using existing equipment, and utilizes an off-the-shelf

methodology which has been proven to be reliable as an indicator of performance.

In order to effectively implement the TRP concept some minor changes will be required in the statistical methods of the NTC. Such changes involve only small recalculations of new statistics utilizing the numbers currently provided by the NTC data gathering structure. Some additional OC inputs are necessary, but these are not burdensome and indeed, border on the trivial. No new instrumented measurements are required for the TRP. The methodology has been created to function within the existing NTC operational framework.

Yet for the effort of recalculating some statistics, certain very tangible benefits will be realized . The TRP will provide the following enhancements to the NTC evaluation structure:

<u>Better reflection of mission accomplishment</u> --The TRP utilizes several accepted Army MOEs [Ref. 21] that will give a clearer picture of actual unit mission accomplishment. The commonly accepted technique of "normalizing" data [Ref. 19:pp. 39-40] permits a more introspective assessment of overall unit performance.

<u>Better understanding of relative performance</u> (without comparison)--Using the TRP, a commander and his staff can gain a truer picture of their organization's relative level of training. The use of a percentile evaluation indicates a degree of relative performance which can easily be used to "key" an item for intensive follow-up training. This system also automatically assists the commander in prioritizing his efforts. Yet, for all these benefits to be gained, no unit comparison or "scoreboard" is necessary; indeed, the methodology almost entirely precludes grading and stays within the ARTEP philosophy.

<u>Allows for greater flexibility in applying standards</u>--Although the NTC Development Plan [Ref. 14: pg. III-7 to

III-11] requires unit performance to be compared to a set of standards, the TRP methodology mandates no fixed standards. Since the commander is doctrinally responsible for fixing standards [Ref. 5: pg.13], the TRP evaluates units based on relative competence only and allows the commander to fix standards at whatever level desired. The TRP also readily lends itself to the calculation of standards when it is deemed appropriate to do so.

Better illustration of complementary weaknesses (cause and effect)--The TRP will provide a more distinct audit trail on areas of performance, i.e. it can better demonstrate linkage between areas of relatively low performance and their effect upon unit performance. Such relationships are often unintelligible when presented in the form of bare numbers and statistics. The format of percentile evaluation allows for greater depth of analysis.

The TRP is a major aid to identifying training weaknesses. It will highlight deficiencies much more efficiently, it is felt, than simple numerical summaries of training events. It will facilitate the commander's own appraisal of his unit's performance and help to focus his subsequent training efforts. This action, in turn, helps to conserve valuable resources and achieve the desired goal--the most efficient, effective training possible under the train-evaluate-train concept, both at the NTC and the unit home station.

A. THE DECILE EVALUATION

An efficient set of measures of effectiveness have been defined that reflect combat training proficiency. It remains to be shown how best to present to the commander the information contained in these measures. A unit can conduct

meaningful remedial training only after it clearly understands where all training deficiencies lie.

At first consideration it would seem that the raw score[9] obtained from the MOE would present all needed information. Indeed, if this MOE really represents a quantitative evaluation of unit performance, can the raw number be improved upon? Certainly, because the raw score only reflects things as they happened at that moment in time when the unit conducted its operation. Other synergistic effects on the instrumented battlefield could have influenced these raw scores on any other given day. This is to say that there is some inherent inaccuracy in every evaluation, no matter how carefully crafted.

> "In the more general sense, the notion of the equality of raw score units clearly violates one's sense of an underlying scale since the raw score scale separations are only the result of the interaction of the particular items that have been put in the test and therefore have no generality." [Ref. 23:pg. 511]

Thus, in order to convey real generality (and hence true performance) these raw scores must be operated on and displayed in the context of relative (normative) performance, which tends to account for within-exercise variations. Some methodology must demonstrate what the raw

---

[9]"Score" as used here refers only to the numbers obtained from the MOEs themselves and is consistent with current evaluation literature. It does not imply a competitive "test" value, and should not be construed to be one.

score of each MOE means when related to the population of heavy battalions in the Army that have trained at the NTC. In this manner a truer unit evaluation can be obtained--one which reflects relative training accomplishment, both Army-wide and between measured areas, or MOEs.

1. Basic Assumptions

Essentially, the procedure being suggested for the TRP is that aggregated totals of unit performance statistics will provide a defacto set of standards against which another unit can be compared. This is not meant to imply "unit comparison", but rather that aggregated statistics over the long haul will provide an adequate set of standards or precisely, norms. This is, in fact, the procedure as envisioned for the NTC by HQ, TRADOC:

> "A secondary objective (of the phase) will be to define baseline (norm based) standards by trend analysis of the data base. Qualitative standards development will be a continuing process until criterion based combat standards for the NTC scenarios are developed. This phase will be continuous. Initial standards will be developed by Aug 80." [Ref. 14:pg. III-10]

There are several assumptions implicit within this method.

a. Normative Performance

The first and key assumption made is that unit performance, on the whole, will reflect an adequate level of mission accomplishment. Certainly, in the absence of numerical standards, it makes no sense to compare a unit

against a set of norms if the normative performance (as reflected in the data base) is inadequate. If, in the data base, most units are failing to accomplish a particular mission, it would be useless to compare another unit against this aggregate. Such a data base only provides the norm for failure.

Therefore, the assumption is made that most units training at the NTC will accomplish their mission. This assumption would seem to be borne out by the fact that historically, in training, units rarely fail utterly in accomplishing their mission. Naturally, degrees of mission accomplishment exist. It is, however, believed that this topic will not prove to be a serious problem in practice.

The second assumption regarding normative performance is that aggregated unit performance will provide an <u>acceptable</u> normative standard. As an example, it is assumed that heavy battalion task forces training at the NTC will not regularly be annihilated in the hasty attack, or any other mission. Certainly if training battalions regularly suffer 85% casualties in order to accomplish a given mission, this would provide an unacceptable, although adequate, standard (i.e. the mission is being accomplished, but at too high a cost). This situation is more likely to

occur than the former. If this assumption proves to be false it can be corrected by "fine tuning" exercises and scenario redesign. On the real battlefield commanders will not regularly assign impossible missions to units which require virtual unit destruction in order to achieve mission success. Therefore, it is assumed that units will be able to accomplish most assigned missions (if properly executed) without sustaining excessive casualties.

b. NTC Procedures

In calculating portions of the TRP it is assumed that standard scenarios of roughly equal difficulty are being employed. Currently, it is planned to draw from a series of pre-packaged scenarios, constructed by mission (see Appendix A), the particular situation to serve as the basis for issuing exercise segment operations orders (OPORD). These scenarios have been developed based on the guidance of the NTC commander [Ref. 24] and are designed to be of relatively equal difficulty for each mission.[10] Several versions of each scenario exist. These versions are designed to play the same scenario (mission) over different terrain. The assumption of equal difficulty is inherently

---

[10]Based on discussions in September of 1981 with MAJ Jim Ireland, UTD-NTC CATRADA, the action officer responsible for developing the scenario package.

tough to validate. However, it is not deemed likely to be a problem, as common military judgement seems capable of producing scenarios of equal difficulty.

It is also assumed that the TRP calculations will be restricted to scenarios of like intensity. Unit commanders may choose the exercise intensity level they desire for their units. These intensity levels are determined by fixed quantitative inputs expressed in terms of OPFOR combat assets. Thus, for example, a hasty attack scenario at intensity level 3 requires certain actions and the presence of certain specific OPFOR assets. As this intensity level is quantifiable, it is assumed that TRP calculations will be restricted to like intensity levels and thereby the TRP calculations will be drawing samples from a homogeneous population.

Lastly, it is assumed that data which is contaminated by machine failure in any significant way will not be incorporated into the TRP statistics.

2. The Use of T-scores

A quick perusal of the MOEs selected for the TRP makes one fact obvious. Although each MOE produces a quantitative number, all are on incommensurate scales or consist of different measured items. Each MOE can, in fact,

be considered to be a random variable possessing its own unique distribution. The problem of comparing these items in order to determine training weaknesses (and subsequently allocating training resources for remedial training) becomes a little like sorting apples, oranges and bananas for "the best color". Each item (MOE) has its own intrinsic scale of measurement. Almost every MOE collects data on a ratio scale, i.e. a scale that possesses a fixed zero point for calibration (see Appendix D), but each scale is somewhat different in terms of measurement units.

Therefore, in order to evaluate relative performance across varying MOEs, some method of transforming this quantitative data becomes necessary. Educational measurement literature describes several scaling methods. Each has its own usefulness, but in particular, the T-score scale is selected here for use with the NTC data.

The main reason for choosing a scaling method lies in the principle that:

> "In no case does a single measurement give us a great deal of information. A single measurement selected from a known distribution tells us much more, for then we can say whether the item picked is exceptionally large or small. In fact, we can say what proportion of cases fall above or below it." [Ref. 25:pg. 325]

Each MOE can, of course, be considered to be a random variable. Thus, one of the easiest methods for conducting

an analysis of relative performance would be to construct a frequency distribution for each MOE, and then compare the measured unit data point to this distribution to obtain a straight percentile evaluation. This is an acceptable procedure and is simple to perform. However, there are disadvantages to this method, and chief among them is:

> "The percentile rank scale is clearly ordinal and, according to most points of view, its units are unequal since they are intended to provide equal proportions of a group, not equal intervals on a scale of ability." [Ref. 23:pg. 515]

One of the intrinsic goals of the TRP is to reflect the performance of a unit across various MOEs on a scale of ability, that is, to help the commander determine which areas most need further attention. Therefore, arranging MOEs on a scale of "equal ability" becomes important. Additionally, if we do not know the distribution of each MOE, it should be considered that:

> "Many times it is difficult to look at an entire distribution, and we should like some way of designating a single score (measurement) so that its value tells at a glance whether it is a comparatively large or small value." [Ref. 25:pg. 325]

Therefore if we wish to use the TRP as a resource and training allocation tool we should obtain evaluations reflecting equal intervals on a scale of ability. The measures of relative performance should be drawn from a known, recognizable distribution with a predetermined mean

and standard deviation for ease of MOE comparison. The T-score method of scaling provides these characteristics.

Finally, once the NTC is in full operation it may be discovered that the instrumented data requires "smoothing". That is, battlefield synergism may cause data to be obtained in a non-continuous manner. A histogram of such data for a particular MOE may reveal "bunches" or gaps in the distribution. This problem can be overcome by employing T-scores. The T-score method can eliminate such synergistic effects or day-to-day variation by "smoothing" the data.[11]

The T-score is a normalizing method. Regardless of the MOE distribution, T-scores are reflected against a normal probability distribution. In this sense it is "distribution free". The T-score method is described as follows [Ref. 25:pg. 325-326]. A baseline data set is defined against which the current evaluated unit is to be related. Given the numerical value (say M) for an MOE, the number is compared to the baseline data set for that MOE. The proportion "p" of cases falling below this value is determined. This "p" value is then used with a cumulative normal distribution table to obtain a "z" value. The "z"

----

[11]This advantage to using T-scores does carry with it an underlying assumption that the data being "smoothed" is, in fact, drawn from a normal distribution.

value corresponds to the point on the normal curve abscissa where the area under the curve to the left of "z" equals "p" percent of the total area under the curve. Now, any variable "T" can be "standardized" by a linear transformation to a designated scale (This is the common method of Z-scores, [Ref. 25:pp. 31-32]). The scale used by the T-score (arbitrarily chosen here) has a mean equal to 50 and a standard deviation equal to 10. This choice will facilitate later comparison efforts and provides the "equal interval of ability" in the form of the standard deviation, which equals 10. The standardized value, (T-50)/10, is then set equal to the "z" value obtained above. The resulting equation,

$$z = (T-50)/10$$

is solved for T which is the desired T-score. [Ref. 25:pg. 325-326]

As an example, if M is such that 90% of the baseline case measures fall below it, then p=.90. Consulting a cumulative normal distribution table yields a z=1.28 for this corresponding p value. This leads to the equation: 1.28 = (T-50)/10. Solving the equation produces T = 62.8.

This score, then, relates unit performance on this single MOE to the baseline data for that MOE. It provides a "normalized" measure that reflects actual unit performance. This measure will facilitate later statistical operations on the data, but for the moment it provides a representation of

relative unit performance on an equal interval scale of ability.

> "It was pointed out earlier that since the properties of the raw score scale, or a linear transformation of the raw score scale, are dependent on the characteristics (difficulties and intercorrelations) of the particular items that happen to have been chosen for the test, it is frequently considered to be advantageous to transform the scale to some other system of units that would be independent of the characteristics of the particular test and, in the sense of a particular operational definition, equally spaced." [Ref. 23:pg. 515]

Therefore, the T-score is used.

3. <u>The Baseline Data</u>

For comparative purposes it is important that the baseline data used as a norm be properly defined. The baseline data essentially establishes the norms against which a training unit will be examined in order to reflect a relative performance evaluation. As such, the items forming this standard reference group must be similar in most important respects to the group being evaluated [Ref. 23:pg. 513].

Once the NTC is operating at full capacity it is planned to rotate 42 battalions per year through the training programs there. Each battalion in the continental US will actually rotate to the NTC for training once every 18 months. Therefore, a total of approximately 63 battalions will be trained at the NTC every 18 months. Because a normative evaluation is planned for the TRP, this population

of 63 previously trained battalions would represent the most current set of "norms" availible to a training unit. In effect, a "moving baseline" is suggested here as the most meaningful normative population.

As a battalion begins training at the NTC, a new baseline data set is calculated. The oldest battalion data set would be dropped from the baseline, while the most current (the 63rd trained battalion, or the last previous battalion data set) would be added. Thus, as the first TRP is calculated for the battalion in training, all MOE data for it would be compared against the previous (most current) 63 battalion data sets on file.

This "moving baseline data" would have several advantages. First, it would preclude inter-battalion comparison and eliminate any possibility of "grading" between battalions in training. Each battalion would be compared against a new set of norms. Thus, relating one TRP to another would render no meaningful comparison. The comparative evaluation provided to the training battalion, however, would still be meaningful as it would relate current unit performance to the _most recent_ performance of unit peers. This would still allow brigade or higher commanders to establish training priorities by using several

TRPs, for although no unit comparison is possible, certainly trends can be identified.

Secondly, such a "moving baseline data" set would allow for gradual training improvement over time. Certainly, as OC's and TAF personnel become more proficient the training value of NTC exercises will increase and likewise unit proficiency will increase. By sequentially dropping and adding data to the baseline, normative "standards" will rise gradually as the salutary effect of NTC training begins to improve Army-wide training readiness. Thus, training units will not suddenly find the norms drastically different from cycle to cycle. Units may then train in confidence knowing their performance will be related to the accomplishments of peer battalions since their last visit. No particular data analysis would therefore be required in order to maintain the data base--the baseline would, in effect, become self-maintaining.

As with any system, however, there are certain problems inherent to the moving baseline data set. Changes in TO&E (organization) or equipment (the substitution of M1 tanks for M48A5 tanks, etc.) are, in effect, changes in the population. Data collected from such units will not be compatible with the data base as originally established and

will incorrectly influence the norms. Thus, new units must be treated separately and new baseline data sets established for them.

At the current time in Phase I of the NTC, it is planned to only instrument armor-heavy battalion task force teams in training at the NTC. Thus, a homogeneous population for the baseline is initially insured. When mechanized infantry task forces are fully instrumented at a later date, their data may prove to be incommensurate with armor-heavy TF data. Since both battalions use the same types of equipment (and differ only in the quantity authorized) the "normalizing" effect of the MOEs chosen for the TRP (see chapter 3) may eliminate this problem. However, this fact must be analyzed and verified before data from both types of units can be comingled in the baseline data set.

These questions can be settled by examining NTC instrumented data as it becomes available. The "moving baseline data" set concept is sound, and can undoubtedly be implemented once these initial questions are resolved.

4. The Calculation of Decile Standing

It is instantly recognizable, however, that a T-score is unacceptable for psychological reasons as an

indicator of unit relative performance. The simple use of the word "scores" is enough to conjure up ghostly visions of the old ATT and ORTT graded systems which the Army has eliminated. Any evaluation method such as is proposed here must scrupulously avoid the appearance, however slight, of a graded result. It must be remembered that the calculated T-score is NOT a "score" or "grade", but rather a convenient method of reflecting and standardizing a number to represent relative performance. Therefore, although the T-score is statistically a very useful result (and can provide a wealth of data to an operations research analyst) the final evaluation must be presented to the unit commander in a different format.

One solution to this problem is the use of decile evaluations. The T-score is a normalized statistic and thus reflects a position on the normal curve. If the area under the normal curve was subdivided into ten equal proportions of area, it would form deciles (see Figure 12). This form of partitioning is very similar to the stanines[12] method used by the Army Air Force in WWII for personnel evaluation.

---

[12]"Stanines" was a general standard score system developed by USAAF psychologists for use in WWII. The plan divided the normal population into nine groups, or "standard nines". The result was a distribution with a mean = 5.0 and a standard deviation = 2.0. [Ref. 26:pg. 128]

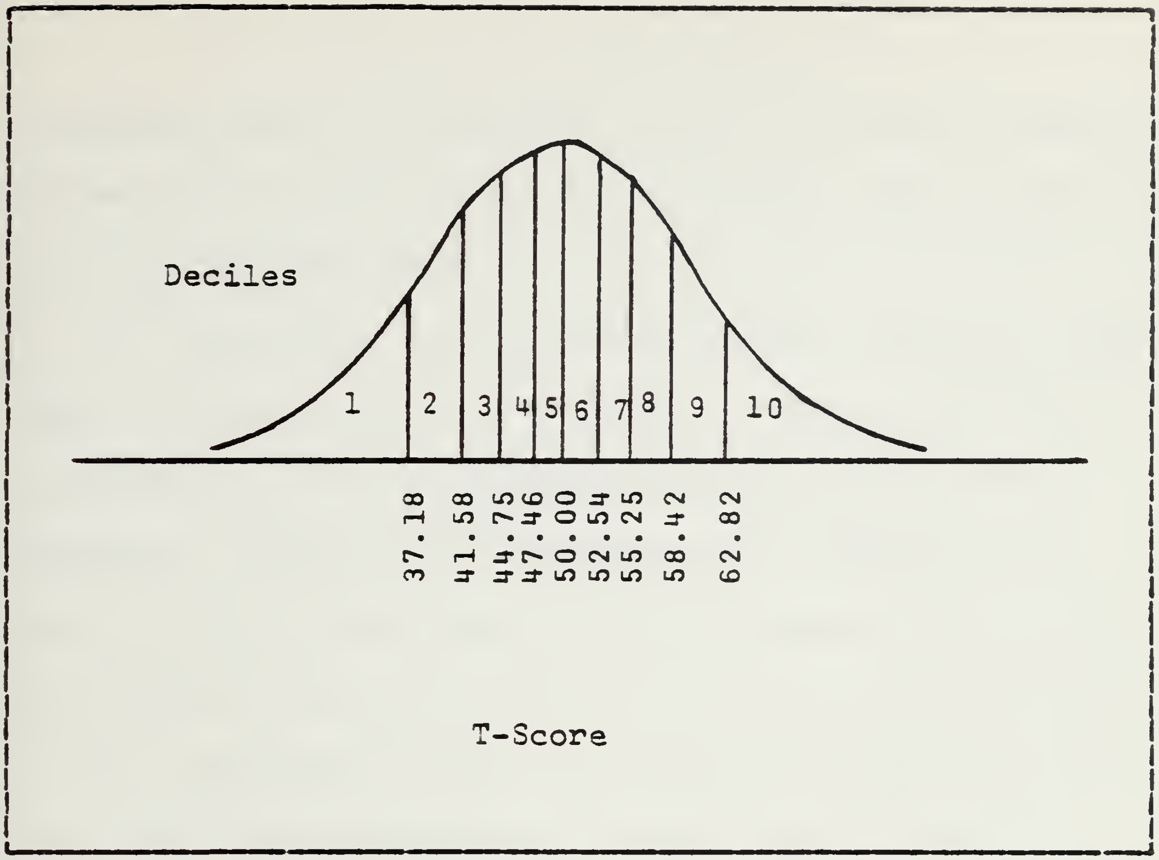


Figure 12: Deciles

Once this decile subdivision has been accomplished, the T-score for each MOE can be placed on the curve into its appropriate decile. This decile evaluation of relative performance can then be provided to a unit commander without any appearance of a score and its associated stigma.

Thus, when several MOEs are displayed in bar graph style side-by-side, comparisons of relative performance in each of the areas becomes easy and almost immediately obvious. An MOE where a unit performs high in a relative

comparison will be represented in one of the upper deciles while an area of relatively low performance will be shown by a bar in the lower deciles.

This system is, of course, somewhat circular. This same information could be conveyed to the unit commander by utilizing the straight percentage calculation initially discussed. However, if the problems of "unsmooth" data occur, or if the data turns out to be sparse,[13] then the T-score method will allow for meaningful statistical analysis of unit performance whereas the percentile evaluation will not. Additional benefits also accrue from the use of T-scores and deciles in the field of statistical analysis (see Chapter 5 for discussion).

Thus, the decile evaluation, coupled with the use of T-scores provides a meaningful, normative evaluation which facilitates analysis to identify areas of training deficiencies. In turn, this helps in resource allocation and highlights the direction of remedial training. A further benefit of this normalized system is that it relates relative performance between the various measured MOEs to the commander.

----

[13]That is, if many units achieve identical numerical totals for the various MOE. In this situation, the data provides no real information about the actual distribution of the MOE.

This is, indeed, a necessary endeavor, for as the educational community dicovered decades ago:

> "An individual's test score acquires meaning when it can be compared with the scores of well-identified groups of people. Manuals for tests provide tables of norms to make it easy to compare individuals and groups. Several systems for deriving more meaningful "standard scores" from raw scores have been widely adopted. All of them reveal the relative status of individuals within a group." [Ref. 26:pg. 1]

Given the great value of the statistically reliable information delivered by the NTC, the Army can do no less. To do otherwise is to discard hard-won, expensive data which, if properly and completely analyzed, could ultimately save lives in combat.

B. THE TRAINING READINESS PROFILE STRUCTURE

The quantitative measures used to provide an objective unit evaluation and the underlying theory for showing relative performance have been given. The next step in this process is to explain how these measures of performance can be displayed for a unit commander. It is important that any presentation be concise yet thorough; that is, address the main issues of training performance and avoid minute details. Numbers that reflect actual as well as relative performance will assist the commander in identifying areas for future training. Any representation should be easy to understand and interpret so it can serve as an aid to the

commander and his staff even after returning to their home station. The remainder of this section will outline a structure for displaying the TRP. For ease in understanding the structure a sample TRP is presented.

1. Organization of the TRP

The TRP will consist of four separate pages. The first page will display the unit performance with respect to overall mission accomplishment. The other pages will reflect the unit performance in its ability to shoot, move, and communicate. No direct correlation exists between the four separate pages. However, specific MOEs that portray low performance on one page may be further explained or supported by MOEs on a different page. An example of this relationship is provided in the sample TRP.

Each page is constructed and presented in the same manner. The top of the page is titled to provide the exercise segment conducted by the unit (attack, defend, etc.). Under this title is the functional area covered by the page: mission accomplishment, shoot, move, or communicate. Following this, there is a bar chart that shows the various MOEs for that functional area along the abscissa. Listed with each MOE are the actual numbers used to compute the value of that MOE. The decile increments are

displayed on the ordinate axis. Above each MOE is a bar that shows the performance of this unit relative to the units considered in the data base. The "shoot" page has three bars for each MOE; each bar depicts a different weapon type.

Interpreting the results shown on the bar graphs will lead to the identification of training weaknesses. In all cases, a bar that fits in the lower decile areas indicates that relative to baseline performance, this unit has a training weakness. Some MOEs are defined so that a small number is better (e.g. all MOE in the communicate area). The TRP is constructed to accomodate this fact. Units with higher numbers will be placed in the lower deciles. The key to understanding these graphs is in realizing that any bar above the 5th decile means this unit performed better than average, relative to the baseline data. Likewise, any bar in the lower deciles reflects the fact that this area needs further training, relative to Army-wide norms.

This system allows the commander, at a glance, to identify areas of performance that contain training weaknesses. The commander can then concentrate his efforts and conduct a more thorough analysis of this area. Using

the detailed data collected by the NTC instrumentation a more specific listing of soldier and unit tasks can be identified for remedial training.

2. Sample TRP

The following pages depict the TRP that could be generated for a typical tank heavy battalion task force training at the NTC. Numbers shown on the sample TRP are ficticious and are not intended to represent any particular unit. The analysis accompanying each page of the TRP is taken from the displayed results and represents a possible interpretation. The general context of the comments concerning unit performance are available from proper interpretation of the TRP. Some specific items in the comments would require more justification. In the case of a real unit being evaluated at the NTC the actual numbers to support these comments would be readily accessible through recorded radio nets, video tapes of activities, or from events recorded by the instrumentation.

Before discussing each TRP page it would be helpful to obtain a clear understanding of what is depicted. The numbers shown below each MOE reflect actual unit performance. As an example, in the area of mission accomplishment (see Figure 15), the friendly force killed 18

of the 30 OPFOR combat vehicles and 239 of the 425 OPFOR personnel that started the exercise segment. The length of the bar above each MOE represents how well this unit performed relative to the units considered in the data base. Since a data base is currently nonexistent the displayed results were determined by assuming a hypothetical data base. Examples of the calculation for relative performance are:

> Killing 18 of 30 OPFOR vehicles (or 60%) places this unit so that 56% of the units in the data base killed a lower percentage. Therefore, p=.56 and from the normal curve table z=.15. This produces a T-score of 51.5 and reference to Figure 15 shows this score to be in the 6th decile as depicted on the TRP.

> Killing 239 of the 425 OPFOR personnel (or 56%) places this unit so that 62% of the units in the data base killed a lower percentage. Therefore, p=.62 and from the normal curve table z=.305. This produces a T-score of 53.05 and reference to Figure 15 shows this score to be in the 7th decile as depicted on the TRP.

A sample TRP follows and is given with a possible interpretation for the various category pages.

a. Mission Accomplishment (Figure 13)

This unit performed well in killing the OPFOR, both vehicles and personnel. This high level of performance in both areas combined to produce an even higher overall OPFOR loss value than expected. From the MOE #8 it seems the unit moved very quickly through the exercise segment. Relative to other units this unit lost a high percentage of friendly vehicles and personnel, but the lowest area is MOE

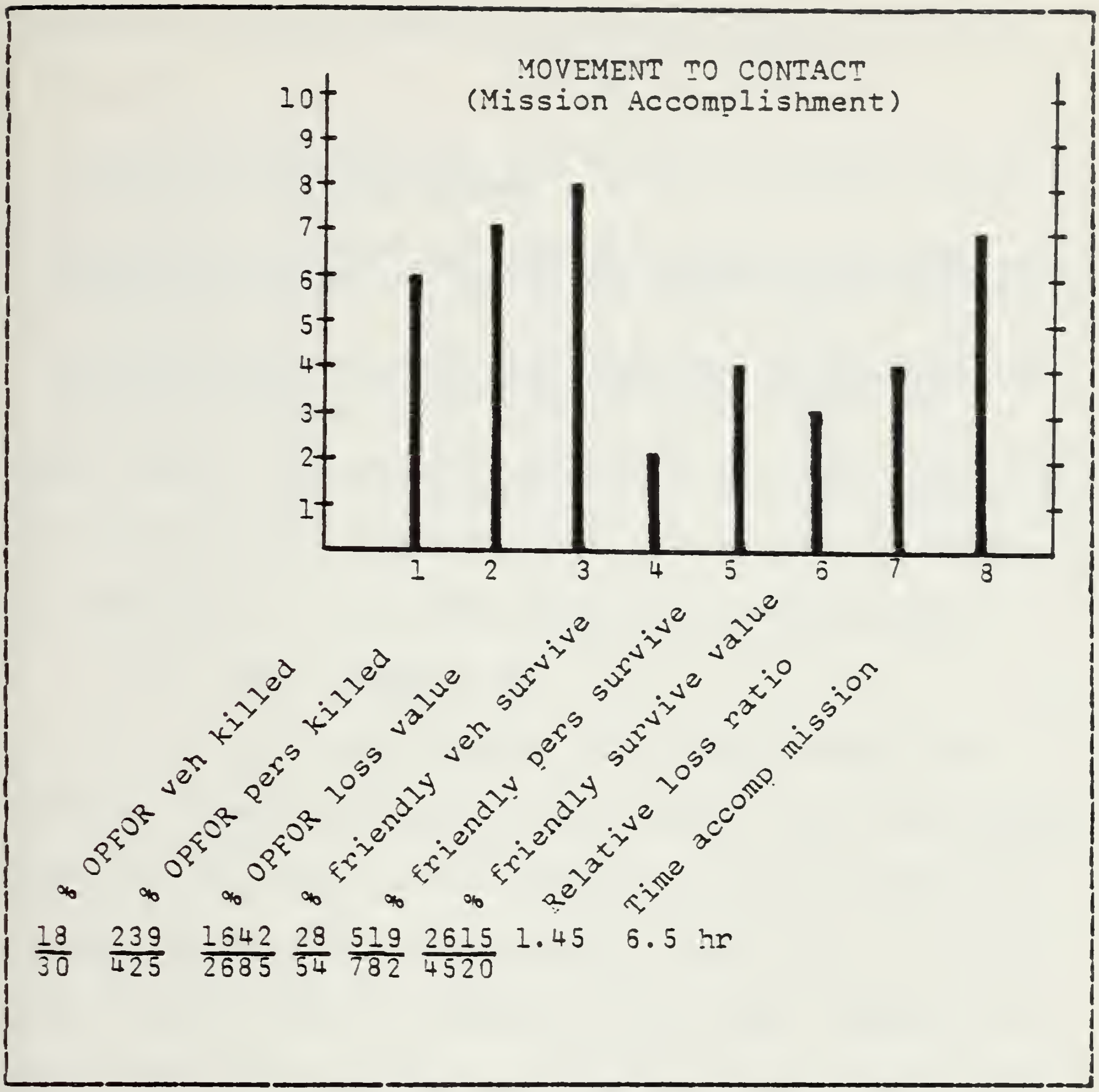


Figure 13: Sample Mission Accomplishment TRP Page

#4, loss of vehicles. Even though the friendly force sustained a large percentage of casualties, MOE #7, the relative loss exchange ratio, was not uncomfortably low since the friendly force was able to inflict heavy losses on the OPFOR. Possible reasons for the areas of relatively low

performance might include (and can be supported with OC comments):

The unit was moving too fast to allow vehicles to use proper movement techniques.

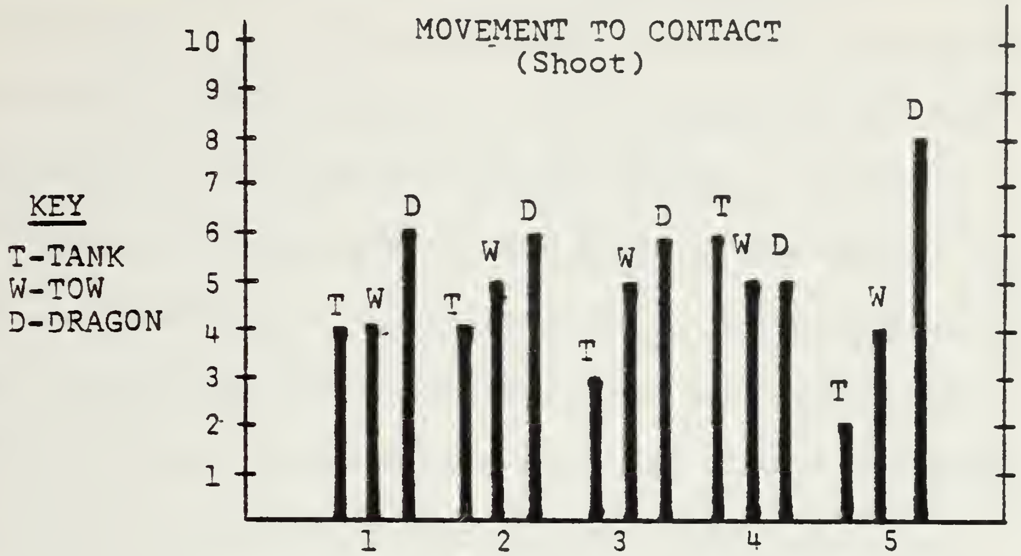
The vehicle drivers and commanders were not trained well enough on the use of terrain for cover and concealment to mask movement.

Since the unit moved so fast there was insufficient time to place forces in overwatch positions to provide protecting fires.

These reasons are not all inclusive of the possibilities and may be incorrect. The other TRP pages need to be examined to determine if any of these possibilities have merit.

b. Shoot (Figure 14)

The tanks and TOWs fired fewer rounds than normally observed for this type of action while the Dragons seemed to fire more. With respect to weapons proficiency, it appears that the tank gunners did not kill or hit many OPFOR targets (MOE #5) relative to the norm. However, the percentage of near misses was relatively high, which means the tank gunners had identified the targets, but were not shooting accurately. Input from the field controllers indicated that since the unit was moving so quickly most tanks were firing on the move. This unit might need additional gunner training to shoot while moving. In the actual data it was found that the tanks killed 10 of the 18



	# rounds fired	Casualties per round	% rounds hit target	% rounds near miss	Wpns fractional kill
T-	41	.25	35	40	56
W-	4	.50	50	25	11
D-	20	.30	32	35	33

Figure 14: Sample Shoot TRP Page

OPFOR vehicles during the exercise. This may seem good, but relative to other units the tanks usually accounted for approximately 70% of the OPFOR vehicles killed. MOE #5 (weapons fractional kill effectiveness) indicates that the Dragons killed many more enemy vehicles than usual while

both the tanks and TOWs were low killers in relative terms. By referring to the battalion operations order and position locations of various players, it was determined that the commander led his movement with the infantry, thereby allowing Dragon gunners many more shots at the OPFOR. This resulted, however, in the friendly force losing several of their own vehicles since friendly tanks were behind the infantry and not in a position to return fire on the OPFOR. The player position plots revealed that TOWs were not employed from available overwatch positions. Rather, they were used with the leading maneuver element and thus their capabilities were limited.

c. Move (Figure 15)

From the mission accomplishment page it is known that the unit finished the exercise segment quickly. This is confirmed by the depicted movement rates. Comparing the two rates, it can be seen that the unit followed a fairly direct route. They only travelled 20 kilometers, and there was 16 kilometers straight line distance from Company A's start point to its designated objective. A map study revealed a different route that was slightly longer but could have provided more rolling terrain for concealment. The leading companies conducted a well coordinated movement

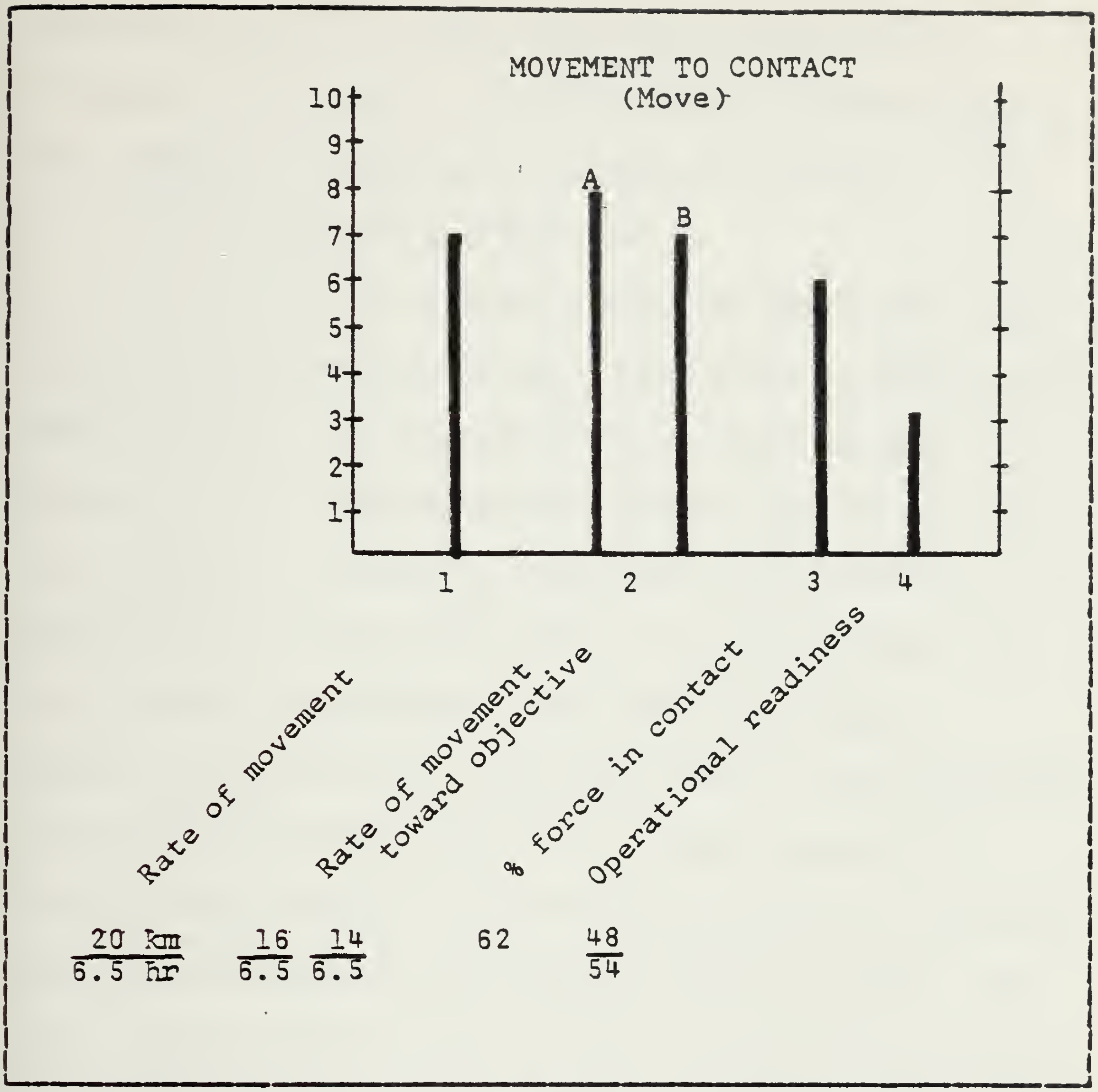


Figure 15: Sample Move TRP Page

and protected each others flanks. Relative to other units a good portion of the friendly force remained in contact with the OPFOR. This pressure might have been the cause of the enemy withdrawing in a fashion that left his vehicles exposed to friendly fire. Operational readiness was

relatively low and a check with observer/controllers indicated that 4 of the 6 vehicles lost for maintenance were lost because of improperly maintained oil levels.

d. Communicate (Figure 16)

Relative to other units, the number and duration of radio transmissions was good. All areas of communication seem to be fine with the exception of radio transmission length, or MOE #4 (transmissions greater than 55 seconds). Each company contributed to this count of 5 messages. By referring to recorded radio nets, it was determined that these lengthy transmissions were caused by company headquarters sending the operations order to their platoons. The OPFOR controllers indicated the OPFOR was able to identify some friendly locations and surmised that increased radio traffic indicated an operation about to begin. The OPFOR reacted quickly and was able to position more forces to the front of the known friendly locations. They did not have ample time to construct prepared positions, however, due to the fast movement of the friendly force. This caused the OPFOR vehicles to leave their fighting positions sooner than they desired, and thus they exposed themselves to friendly fire.

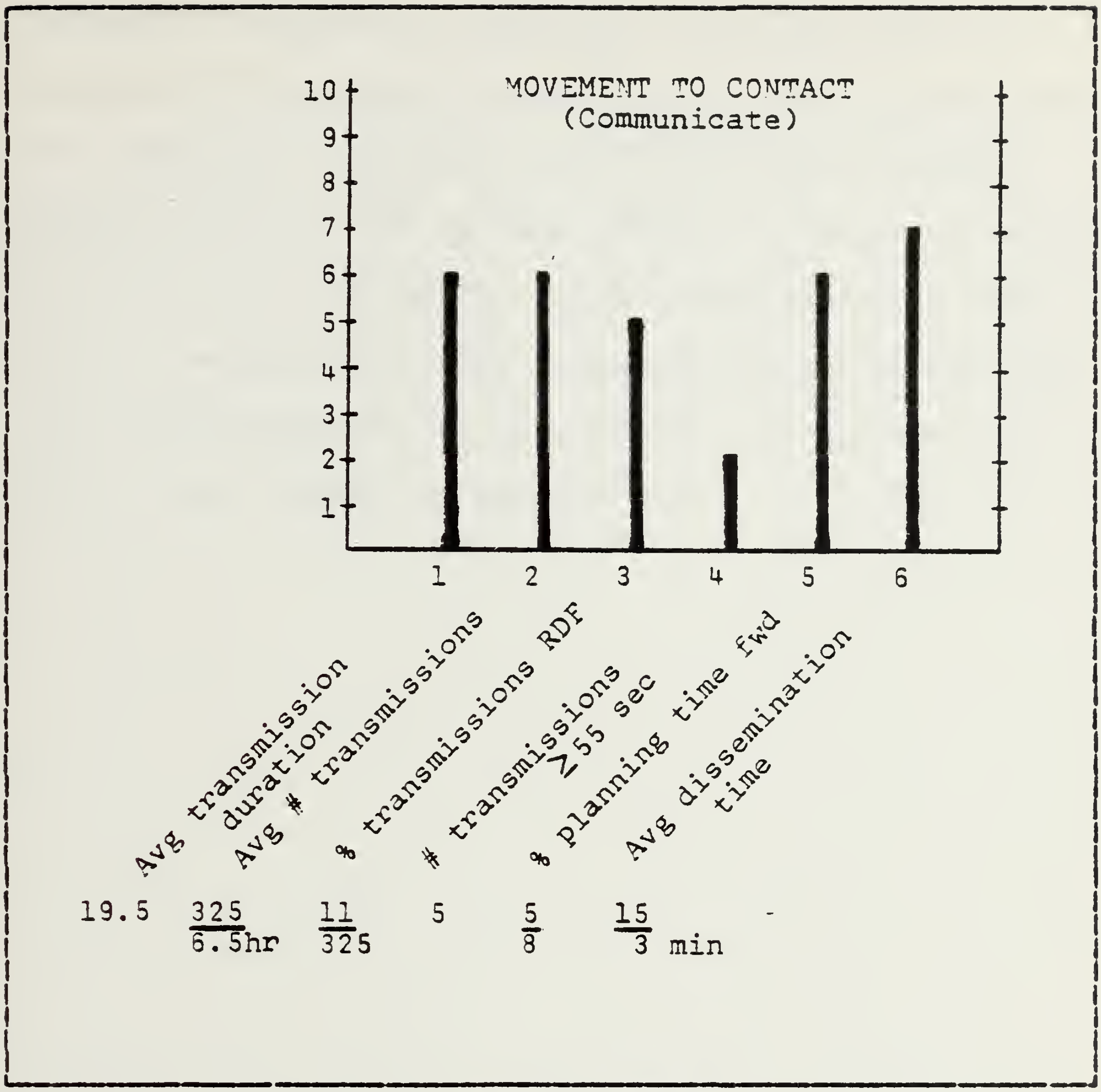


Figure 16: Sample Commo TRP Page

The foregoing example TRP and its accompanying analysis is intended to demonstrate how the TRP could be utilized to assist the battalion commander in reaping the maximum training benefit from the NTC. The TRP is envisioned as being incorporated in the after action review and is not

intended to replace the formal AAR. The TRP should be presented and explained in detail at the AAR to insure that unit commanders can best utilize the document for home station training. The intended primary recipient of the TRP is the battalion commander and his subordinates, as these training managers are directly responsible for troops--and the accomplishment of the real mission: the application of combat power against the enemy in war.

V. UTILIZATION AND EXPANSION OF THE TRP

A key point stressed throughout this paper is that the NTC exists for the purpose of training heavy battalion task forces. In the train-evaluate-train cycle, the evaluation must be accurately directed at training weaknesses. The unit can then use this evaluation to guide its future training and resource allocation. It has already been demonstrated how the TRP helps to identify the areas of training deficiency. The issues of concern now focus on the question of how will the TRP aid the commander as training levels improve over time or as new areas of performance are considered in the evaluation process. Acceptable levels of performance can be established and checks should be made to insure the unit evaluation is being based on meaningful measures of effectiveness. Techniques for expanding and using the TRP to address these issues will be covered.

Although it has not been emphasized so far, the NTC has a secondary mission to:

> "Gather information to help improve doctrine, tactics, training system, equipment, and procedures. This information also assists the Army in relating resources to readiness." [Ref. 13:pg. 1]

It is proposed to use a trend analysis of the NTC data for assessing current Army efforts. The TRP presents an ideal

structure for summarizing relative unit performance trends. This allows for easy comparison of performance for different years or any desired time frame. A brief discussion of the TRP usefulness in this area will be provided.

A. ESTABLISHING COMBAT PERFORMANCE STANDARDS

Despite the inherent advantages and usefulness of the normative evaluation, this type of evaluation has one major deficiency: it does not follow the pattern mandated for Army training in the performance oriented training (POT) system.

The POT system requires that an element's actions be compared against a fixed standard, and that training will continue until the standard is achieved. It has previously been pointed out that although this process is fairly easy to follow for the individual soldier, it becomes increasingly difficult to do as the element grows larger in size (see Figure 4). In fact, one reason given in support of the TRP was that such objective standards may be impossible to set for the synergistic action of battalion level combat. It would be difficult to say that a battalion which siezed the objective, captured 85% of an enemy force, and killed 10% had failed to accomplish their mission. However, this could be the absurd judgement made if the

"minimum enemy casualty" standard for this particular exercise was, say, 30%. The point here is that there may be an infinite set of combinations of standards which would lead to victory. There is reason to believe, however, that a few fixed standards, if applied with common sense, can be developed to reflect minimum acceptable performance (MAP) levels. The TRP can be singularly useful in doing this as it begins with a finite set of "combinations of standards" in the baseline data.

1. Determining Standards

The problem of determining standards for performance at the NTC has already been considered by Headquarters, US Army Training and Doctrine Command (TRADOC). Indeed, a model has been proposed for this process (see Figure 17). The initial analysis of this model has been completed, and only the qualitative application--the actual setting of standards--remains to be done. The next and last phase planned is: "Efforts...will be to integrate the initial analysis with the realities of actual operations." [Ref. 14:pg. III-10]. A great deal of work has gone into the top-down analysis for subtasking the combat process (see Chapter 2 and Appendix C). Therefore, all that remains to be done

FUNCTION	OUTPUT EXAMPLE
Mission/task analysis	Movement to contact
Organization level performing	Platoon
Identification of level of evaluation (activity)	Execution
Identification of EEA	Tactics
Development of MOE	Move under fire
Development of MOP	Percent friendly casualties
Qualitative application	No more than 20% acceptable initial contact

Figure 17: NTC Standards Development Plan

is the integration of reality and analysis. It is in this area that the TRP can provide real assistance.

There are basically three methods which can be used to set "fixed" numerical standards. The first method consists of the normative approach already proposed. Using the TRP, the baseline data set can provide an average value for every MOE used in the evaluation. If all units are accomplishing their assigned missions, and subjective[14] analysis reveals that pertinent indicator MOE values are within the desired range (casualties, percent enemy killed, rounds expended, etc.), then pure MOE average values can be used as a standard. Basically, this amounts to a validation of both the normative TRP system and the current Army organization, doctrine, and training.

If, on the other hand, the average values of MOEs prove to be unacceptable, then a second system can be employed. A subjective search of the baseline data could be undertaken to find those units whose performance in all or most of the MOEs was above a fixed decile level, say, the 7th decile as an example. An average of the MOE values

---

[14]These judgements must by nature be subjective, and the perogative of the Army's senior (general) officers. If the NTC process is accepted as being representative of real combat, then such decisions as what constitutes "acceptable" casualties must be made at the highest policy setting levels.

recorded for these units could be taken and these numbers employed as a standard for the various MOEs. This method rests upon the theory that well-trained units perform most operations well, and any unit which is evaluated at or above the 7th decile in most MOEs could be considered well-trained. Units with high morale, esprit, and cohesion generally perform at a high level most of the time. The performance of a less well-trained unit would tend to be more sporadic by MOE.

A third method for assessing standards from the TRP would be a combination of the previous two. For some MOEs the straight baseline average value might be acceptable and for others, the "search out the best units" technique could apply.

Certainly, in employing any of these techniques, the subjective assessment of senior Army officers must also be injected. As an example, the NTC Commander might feel that units are failing to aggressively conduct operations and are overall suffering an unacceptably high level of casualties. The standards for casualties sustained could be arbitrarily raised to establish a training goal for units to achieve. This is well within the perogative, and indeed is the responsibility, of the NTC Commander or other more senior

commanders. In fact, the TRP system possesses such flexibility that it would be possible for large unit commanders (perhaps division) to set their own unit standards for training. This standard could also be reflected on the TRP to provide one additional piece of information for training assessment to the training battalions. A method for displaying these standards will be provided in the following section.

It should be apparent from the above discussion that standards for unit performance can and should be established. Whatever methodology is employed, the TRP structure and concept provide a useful asset in determining the numbers. Once standards are set, they can be revised as deemed necessary.

2. The Presentation of Standards on the TRP

It has already been demonstrated how the TRP will assist a commander in identifying training weaknesses. This is a tremendous asset to the commander when he develops the unit training calendar and allocates vital resources. The benefit of the TRP relative performance indicator has been discussed. It provides the commander with an objective assessment depicting which areas of his unit need the most

training as well as his unit performance relative to other units who trained at the NTC.

There is another important piece of information that can be displayed on the TRP. Once the minimum acceptable performance (MAP) levels have been established, as outlined in the previous section, they should be reflected on the TRP. A commander is certainly concerned about the relative performance of his unit for various training areas, but it is not sufficient to perform at an upper decile level if the minimum standard has not been achieved. Likewise, a unit may be assessed at a low decile level of performance, but if the minimum standard was still accomplished the commander can concentrate training efforts in areas where the MAP was not achieved. This will serve as an additional aid to the commander in prioritizing his allocation of limited resources for training.

Portraying this standard can be accomplished with minimal effort. It can be reflected in a manner to make TRP interpretation easier and more meaningful. As all MOEs are defined separately, an independant MAP should be established for each MOE. Given a number that represents the MAP for a specified MOE, it will be compared against the current data base to find the percentage of units that did not perform at

or above this level. This percentage becomes the "p-value" used to convert the MAP number to a T-score. A horizontal dash will be drawn in the MOE column at the decile rating associated with the calculated T-score.

To provide a better understanding of the MAP concept, the communicate TRP page from the sample in chapter 3 has been reproduced below. Hypothetical standards have been established and are depicted on the TRP. A calculation to determine the MAP plotting level has been shown. Finally, interpretive comments have been provided to demonstrate how this additional item of information, the MAP, can aid a commander in his training evaluation assessment.

> For this type unit in a movement to contact mission, the current standard is to attain an average dissemination time of 10 minutes. According to the data base 19% of the units failed to achieve this standard. Therfore, p=.19 and from the Normal curve table z=-.878. This produces a T-score of 41.22 and reference to Figure 15 shows this score to be in the second decile as depicted on the TRP (see Figure 18).

Additional interpretations of the TRP page could be as follows (see Figure 18):

Even though the unit did relatively well in MOE #1 (average transmission duration) the MAP was not achieved. One other area which fell below the MAP was MOE #4 (number of transmissions exceeding 55 seconds duration). If the commander concentrates efforts in training radio operators

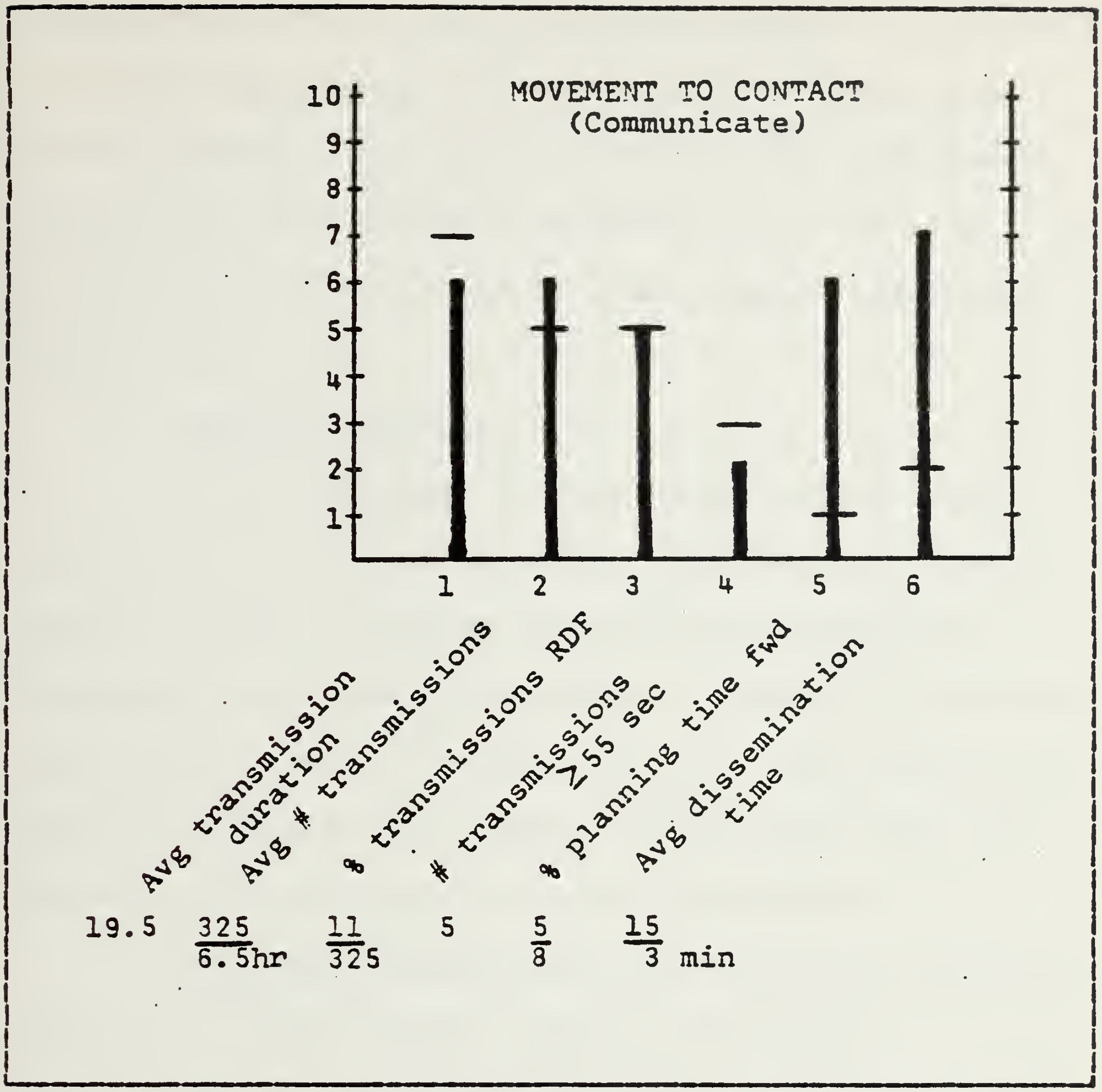


Figure 18: Sample Communicate TRP Page With MAPs

to shorten messages the radio transmission length will decrease in both areas. This should be priority one for communication training.[15] The unit performed relatively well

---

[15]The commander must still assess a training resource tradeoff between major areas, weapons proficiency and communications for example. This decision should be aided by the mission accomplishment TRP page and the overall exercise evaluation provided by the TAF.

in MOE #2 and #3. It should be noted, however, that the MAP was barely achieved and this training area should receive second priority. In the other categories being evaluated the unit not only performed relatively well, but also exceeded the MAP. These areas do not require additional training resources.[16]

3. Revising Standards

The establishment of MAPs and the benefit they provide has already been discussed. To serve this useful purpose, however, requires that qualitative standards development be a continuing process. Revisions to the MAPs are necessary for two reasons: the Army force structure or operating procedure may change, and hopefully, units will become trained to a higher level of proficiency.

Army combat developments activities are constantly analyzing weapons systems, tactics, doctrine, etc. as used by our forces, to determine if they can be improved. When improvements to combat effectiveness are possible the procurement of a new weapon system or an appropriate change to tactical employment may be initiated. The improved

---

[16]Althogh not depicted on this TRP page there might be an occassion when a unit performance is relatively low, but still achieves the MAP. Areas where the MAP was not achieved should receive further training emphasis before an area where the MAP was achieved, even if relative performance was low.

effectiveness usually means a better performance capability of the friendly force.[17] By natural extension of reasoning, this should mean that the standard for unit performance should be raised.

The purpose of the NTC is to provide a location where heavy battalion task forces can conduct realistic training. As units train at the NTC they will be provided with an assessment that outlines their training weaknesses. If commanders use this asset to develop training plans they will certainly allocate resources to the areas that require the most improvement to attain the MAP. This will hopefully produce units that are more combat proficient in these areas (see Figure 7). As the level of training improves the standard for unit performance may be raised.

B. CHANGING THE BASIS OF EVALUATION--SELECTION OF NEW MOES

An assessment of unit performance is based on the areas that are observed and/or measured. The assessment is only meaningful when it remains within the limits of the areas observed and the measurements taken. A unit evaluation compiled from the TRP is derived from the analytical base of

----

[17]Cost and operational effectiveness analyses conducted by the Army are a prime example of this. Before a new system is developed it must be determined that the additional effectiveness provided to the force by the system is worth the cost required.

the TRP, that is, the defined MOEs. The complete unit assessment is supported by other collected data, but the MOEs provide the main direction for an evaluation. Therefore, any meaningful assessment of unit performance should be limited to the areas addressed by the selected MOEs.

When the MOEs were presented in Chapter 3 it was noted that there was no unique set of measures to use for an evaluation. The current TRP graphs have been outlined for the purpose of assessing unit performance in the areas of mission accomplishment, shoot, move, and communicate. If specific functions in these areas are to be asessed then the MOEs can be redefined to measure selected items of performance in greater detail, i.e., through more detailed subtask analysis. As the data collection capability of the NTC expands, the span of evaluated functional areas can be increased. Under the present TRP structure the subjective input of the controllers is not directly incorporated into the unit assessment, but this can be done. Each of the enhancements mentioned here will be discussed in the following sections.

1. Evaluating Current MOEs

The MOEs defined for use with the TRP should be viewed as an initial starting base. They were selected because collectively they provide a measure of overall unit tactical performance. These MOEs should assist in establishing meaningful standards for acceptable levels of performance.

As the level of training improves across the Army, more and more units will achieve the MAP level for certain MOEs. A "significance analysis" of those MOEs should be conducted on a routine basis to determine when all or most units are regularly attaining the desired standard. When most units have attained the MAP level for an MOE, the MOE tends to lose importance or significance in the evaluation process, i.e. it does not provide an indicator of training deficiencies. If this situation occurs, the MOE should be considered for elimination from the TRP. This change can be accomplished from cycle to cycle without regard for unit scheduling, as there is no comparative value to the TRP. Changing MOEs from one battalion to the next is immaterial to the TRP concept as the TRP is not intended to be used for inter-battalion comparison, but rather as an asset for identifying training weaknesses. If the units have achieved

the MAP level then their training resources should be directed to a different area.

As some MOEs lose importance and are dropped from the TRP, it would become possible to incorporate new measures without the size of the TRP becoming unwieldy. Additional MOEs might focus on providing a more in-depth measure of particular performance areas. If desired, the detailed top-down analysis of unit tasks conducted for the NTC could be used to provide candidates for new TRP MOEs. Any MOE proposed as an expansion to the TRP should be checked for the following criteria:

It must possess an operational definition that can be measured.

It must provide a measure for a relevant area of unit performance.

The information gained from it should not already be provided by some other MOE.

Provided these conditions are satisfied, the MOE could be added to the TRP structure on an appropriate page.

The volume of data collected at the NTC will grow quickly as the number of training units increases. Once a data base has been established an analyst can conduct a correlation analysis on the various MOEs. In some cases, two or more MOEs may vary in a related fashion, i. e. under a given set of conditions they change in a proportional

manner. When these relationships are found, the extent of the correlation can be determined. If the degree of correlation proves to be significant, then some of the MOEs may be dropped from the TRP without losing valuable information. This is possible because the remaining MOEs can be used to estimate the performance in the area of the eliminated MOE.

The method of correlation analysis goes beyond comparing selected MOEs. In the sample TRP several assessments were made that required substantiation from sources other than the TRP. The process of relating these other data items to the TRP is referred to as correlation.[18] In the following section it is explained how subjective controller inputs and other measureable areas can be structured into measurements for MOEs. This will assist in the process of correlating all available data to produce a more thorough evaluation of unit performance.

2. Scaling OC Inputs

At the present time the OC inputs determined at the NTC are of little statistical value (although of great value

----

[18] When the correlation involves subjective measures of only an ordinal or nominal scale, nonparametric tests of correlation are available. See Siegel [Ref. 27] for a complete discussion of the Spearman and Kendall correlation coefficients used for this purpose.

for subjective evaluations). Certainly the OC plays a large role in the actual operation of the NTC, for their evaluations provide not only the standardized "professional evaluator" feedback, but also provide the only observations of battalion command group operations. This in itself makes them indispensible, for the actions of the battalion command group can often determine victory or failure regardless of troop training levels.

The battalion command group observation plan for the NTC is well thought out and calls for the subjective evaluation of thirty different MOEs by the OC [Ref. 28]. These MOEs cover the full range of battalion command group operations. It is planned to record the OC observations from this plan in the CIS data base for cataloging and use during AARs. Edited lists of these comments are to be provided to training battalions in the take home package (THP).

Two further evaluations are also planned for the battalion command group and staff. The first is the Battalion Command Group/Staff Evaluation form [Ref. 29:Incl 1,pg. 26]. This evaluation incorporates MOEs from the original plan [Ref. 28], but requires each OC to rank the battalion command group performance on a scale from one to

ten. The purpose in doing so is mainly geared toward assisting the NTC in assessing trends in battalion training and OC standardization. As the concept paper states:

> "There is no intention to use these values as part of the feedback to the training unit. Analysis of these ratings should not be used to compare one unit against another, or for the purpose of assigning a level of performance to the battalion." [Ref. 28]

These ratings were designed to extract the following information:

> "a. OC reliability by comparison of the assigned ratings and the objective measures from the instrumentation.
>
> b. Correlation of a selected MOE or MOP to overall effectiveness.
>
> c. Training trends across all visiting battalions. Once a trend is identified then a detailed testing plan can be developed for more detailed study.
>
> d. Measure the effectiveness of NTC command and control training of unit by graphing the ratings over time." [Ref. 28]

However, there is a great deal more information contained in this evaluation if it can be gleaned from the raw data. One of the key components of the statistically significant evaluation is the presence of trained evaluators capable of rendering a standardized judgement. The NTC Command Group Evaluation Plan [Ref. 28] is reluctant to ascribe reliability to these numerical ratings, as indeed they are categorical judgements made by human beings (not machines) and hence, subject to a variance in accuracy.

However, there is inherent value in the judgements of human beings. Although one of the major weaknesses of the ARTEP is its reliance on human subjective evaluation, this does not mean there is no worth to such judgements. Some hold the opinion that the lack of standardization in training and scenario leads to inaccuracy in ARTEP evaluations more than the internal variance of human judgements does. If we assume that a judges "feelings" about the scale value of an instance (here the scale is 1 to 10) is a normally distributed random variable, then it possesses some mean and variance. This fact can be exploited to follow a technique for constructing an interval scale using data obtained from categorical judgements.

Fortunately, such a technique exists. In an unpublished paper entitled, On Constructing Interval Scales Using Data Resulting From Categorical Judgements, (Naval Postgraduate School, 1981) [Ref. 30], Professeor Glenn F. Lindsay outlines in detail a method by which categorical judgements exactly like the OC 1 to 10 scale can be connected to an interval scale (see Appendix D). This method can be used for any number of judges, and is suitable for machine computation. The method as described [Ref. 30:pp. 1-15] literally constructs estimates of the mean and

variance of the particular "feelings" (ratings) of judges and ultimately, through the assumption that these feelings are normally distributed, produces an estimate of the MOEs value on an interval scale. The method also produces estimates of the upper and lower bounds for each category, i.e. the categories 1 to 10.

The end result of this method is a particular value for this given MOE (drawn from the Battalion Command Group/Staff Evaluation form, Figure 19) arrayed on an interval scale which shows the relative interval width of each category (1 to 10). These category widths are different for each MOE as each MOE is evaluated as a separate entity. In applying this concept to MOEs evaluated by OCs at the NTC, the value given as a final ranking for a particular battalion must be transformed in some manner to be used on the TRP.

What is obtained as an end result of applying Lindsay's method is, for each MOE on the Battalion Command Group/Staff Evaluation form, a value for each battalion in the baseline data located on an interval scale of varying category width. This information provides all that is necessary for the TRP. As the battalions in the baseline data set have been ranked, in effect, on this interval scale

Use the scale shown below to rate the overall performance of the Command Group and Staff as a unit on each item. Determine the quality of performance based on your observations and use the scale that best fits your assessment. Specific examples to support the ratings should be recorded on the evaluation plan. These comments will provide input for after action reviews and preparation of the "Take Home Package".

MARGINAL			ADEQUATE				EXCELLENT		
1	2	3	4	5	6	7	8	9	10
Low		High	Low			High	Low		High

___ 1. Develop plan based on mission.

___1.1 Analyze missions.

___1.2 Identify critical combat information and intelligence.

___1.3 Analyze friendly capabilities.

___1.4 Select routes/zones to objective.

___1.5 Select battle positions.

___1.6 Plan and coordinate fire support.

___1.7 Plan for terrain modification and breaching operations.

___2. Prepare and organize the battlefield.

___2.1 Determine critical place and select a course of action.

___2.2 Select control measures.

___2.3 Communicate/coordinate plans and orders.

___2.4 Prepare logistical estimates and plans.

___2.5 Prepare personnel estimates and plans.

___3. Intelligence preparation of the battlefield.

___3.1 Gather critical combat information and intelligence.

___3.2 Analyze OPFOR.

___3.3 Disseminate critical information and intelligence.

Figure 19: Battalion Command Group/Staff Evaluation Form

for this MOE, the proportion of rankings above and below the value assigned to the current training battalion can be calculated. This provides the p-value necessary to calculate a T-score, and results in a decile evaluation for the TRP.

The current Battalion Command Group/Staff Evaluation form contains eight MOEs with each MOE possessing from two to seven MOPs. This scaling procedure could be applied to each MOP individually, and if desired an entire TRP could be generated to reflect the OC evaluations. This same procedure applies without modification to the Process Assessment form [Ref. 29:Incl 2], the second evaluation planned for the battalion command group/staff. This was designed by the Army Organizational Effectiveness Center and School (OECS) to evaluate the within-group (Command Group/Staff) process of coping with missions and external stimuli [Ref. 28]. Obviously, if this wealth of information is not desired the OC input could be scaled down by calculating TRP values for only selected MOEs and including them in the previously outlined TRP.

This method must, of course, be used carefully, as it is much more subjective by nature than the machine gathered data. The basic assumptions of the Lindsay method

need to be studied in conjunction with real OC input data to determine questions of the reliability of OC input, and whether the method is truly appropriate. These questions cannot be examined until real OC-obtained data is available. The method is offered here as a suggestion for expansion along a path that could prove to be very profitable for the Army. It is not inconceivable that most battalion performance may be highly correlated to battalion command group processes. Thus, investigation with the TRP could yield the real key to determining how best to train battalions in a cost effective manner. It may be possible to determine the empirical relationship between command group performance and unit proficiency. If this occurs, real cost effectiveness in training could be achieved, as it is certainly cheaper to train a group of several officers than a unit of several hundred soldiers.

3. Expansion of Evaluation Capability

Certainly the most fruitful areas of expansion for the TRP lie in the direction of those EEAs identified as important in the top-down anlysis, but not yet incorporated into the NTC instrumented environment. The combat service support (CSS), nuclear, biological, and chemical (NBC), command, control, and communication (C^3) and intelligence/

counter intelligence EEAs are important on the battlefield. Any increased reporting procedures which can help to properly evaluate these functions will ultimately improve the overall evaluation of a unit. These functional areas need to be further analyzed so that the processes involved in these areas can be identified. Once the processes are enumerated, quantifiable MOEs can be defined for the basis of an evaluation.

The first area, CSS, is most amenable to further instrumentation and should be equipped at the earliest possible time. If CSS vehicles could be instrumented to reflect even just position, a wealth of CSS data could be obtained. Instrumenting fuel tanker vehicles to reflect flow rates, and supply vehicles to indicate type of load, i.e. ammunition, food, etc., could greatly improve CSS measurement. These improvements would allow the determination of MOEs directly related to CSS evaluation and thus this topic could be included in the TRP.

NBC evaluation will, for the forseeable future, be relatively uninstrumentable. OC input could be used to assess NBC casualties if real gas (non-lethal training agents) were employed against training units. Such casualty assessments could be reflected in the mission accomplishment

portion of the TRP as an adjunct to percent friendly personnel survived (PFPS) and plotted on the TRP adjacent to PFPS in a fashion similar to the multiple weapons MOEs. This crude assessment would be flawed, however, since many casualties would undoubtedly escape controller detection, but would at least provide a rough determination of NBC defensive capability.

C^3 and intelligence/counter intelligence remain as two areas difficult to evaluate. Little can be done at the moment to instrument such evaluations since these are primarily "people" assessments. The OC may already possess an effective tool for evaluating these areas. The Battalion Command Group/Staff Evaluation form contains some C^3 and intelligence considerations. These could be expanded by defining additional MOEs [Ref. 29]. Using the technique outlined in section 2 above, it could be possible to assemble TRP pages containing evaluations for these areas. This subjective evaluation is at the mercy of the limitations pointed out previously, but at least it offers a "rough cut" at determining a true evaluation of their contribution to unit performance. Thus, C^3 and intelligence MOEs could be reflected as appropriate on the TRP.

These suggested expansions of the NTC capability are by no means complete. Competent analysis of NTC operations after the instrumented battlefield is fully operational will no doubt reveal other methods for evaluation. These suggestions offer a beginning, however, based on the TRP methodology. The rapid and complete dissemination of evaluation results to the training unit can only increase training proficiency and aid the commander in allocating limited training resources.

C. STATISTICAL ANALYSIS

The NTC exists to train the heavy battalion task forces, and entertains other functions only when they do not interfere with unit training. This is absolutely correct as a policy. However, the wealth of data gathered at the NTC cannot be allowed to lie fallow. The establishing authority recognizes this fact, and states that:

> "A trend analysis of data will be useful in assessing current doctrine and tactics and evaluating training development efforts...Testing of new concepts, equipment, or systems will be permitted at the NTC when it does not interfere with the training or the training environment." [Ref. 13:pg. 2]

A great deal can be accomplished with the NTC data in its present format, but even more can be done with the instrumented data if it is drawn from the TRP structure. There is much that can be learned about the combat process

in this manner. Only a few simple statistical analysis techniques are needed to begin reaping rewards from this available data. A basic approach is suggested here. This is by no means an all-inclusive listing of available statistical procedures, and should be viewed as merely a starting point for analysis purposes. Certainly the shape and form of real NTC data, once it is obtained, will determine the appropriate statistical inference techniques.

1. The $\bar{X}$ Statistic

One of the most useful statistics in probability theory is $\bar{X}$, the average of a data sample. If we regard the MOE as a random variable, then at any given moment the baseline data set contains 63 historical sample points for a given MOE. The Central Limit Theorem [Ref. 31:pp. 255-260] then allows one to use $\bar{X}$ as an estimate of the population mean, μ, as long as we are sampling from a normal population. This presents a minor difficulty, as the assumption that the MOE distributions are normal has not been made explicitly so far. This situation will require the judgement of a competent analyst. Before performing any operations on $\bar{X}$, the data set must be transformed to normal (if not already normal). Having done this, a number of operations can be performed with $\bar{X}$.

Since $\bar{X}$ is the mean of a random sample of size n=63, drawn from a normal population with mean = μ and variance = σ^2 its ($\bar{X}$'s) sampling distribution is also normal with mean = μ and variance σ^2/n. It can be demonstrated that for the normal population, $\bar{X}$ is the maximum likelihood estimator of the population mean. Thus $\bar{X}$ yields some concrete information about the parameters of the baseline data set population. This population is, of course, the set of heavy battalions in the continental US training at the NTC. Hence, $\bar{X}$ reflects the attributes of the Army as a whole.

This fact can be useful for trend analysis. The NTC will certainly have a salutary effect on training levels throughout the Army. A time plot of $\bar{X}$, using the baseline data set, should reveal Army-wide progress trends for each MOE. This information would be particularly useful in assessing the effectiveness of NTC training over time and in future justification of the NTC operation (budget). To ascertain more accurately the actual significance of these $\bar{X}$ values, confidence intervals about $\bar{X}$ can be calculated for given levels of significance using standard statistical practices. [Ref. 31:pp. 342-346]

The calculation of a confidence interval about $\bar{X}$ might also yield other useful information, as it would

reflect the "spread" of the $\overline{X}$ statistic considered to be significant. An appraisal of the width of that confidence interval might give some subjective idea about how precise a reflection of training particular MOEs are. If the confidence interval is large, it may be due to a fundamental difficulty in the MOE; that is, the MOE may not be an accurate reflection of the quantity being measured.

Hypothesis testing of $\overline{X}$ could also be conducted to assist in trend anaysis. Using one-tailed tests, it would be possible to determine the direction of change in the mean from period to period (quarter to quarter, month to month, or whatever report interval is desired). That is, given a period 1 preceeding a later period 2, a null hypothesis that the mean level of performance remained the same between periods ($\mu_1 = \mu_2$) could be tested against an alternate hypothesis that the level of training improved ($\mu_1 < \mu_2$). If the statistical test allows the acceptance of the alternate hypothesis that the mean levels of performance within the random sample have risen, i.e. $\mu_2 > \mu_1$, then it can be assumed that the MOE averages are rising and the level of combat proficiency in the Army is improving.

This is overall an exciting concept, because for the first time Army training progress could become verifiable,

and the increase in large unit skills documented. This is in no way possible now with the ARTEP system of evaluation. The ARTEP system has long contributed to the abuse of readiness reports (readiness condition levels) due to its subjective nature. Using the TRP, it may now become possible to accurately reflect Army-wide readiness levels.

As this methodology for $\bar{X}$ analysis is keyed to the baseline data set, it would be impossible for the moment to determine individual unit readiness from the NTC. However, trend anaysis as previously described may yield a linkage between MOE values and unit readiness, and thus readiness conditons could be evaluated by the NTC results. This is probably not advisable, however, as this converts any NTC exercise into a defacto test. Thus, in order to preserve the NTC as a training ground, any assessment of readiness levels should be restricted to an Army-wide basis.

This problem not withstanding, if desired readiness goals can be linked MOE $\bar{X}$ values by judicious analysis (or by arbitrary decision based on senior officer combat experience), a powerful assessment tool has been created for the Army Staff. By observing the trends and viewing the progress of the Army as reflected in the baseline data set, the Chief of Staff and other senior Army leaders could gain

a much more concrete and reliable assessment of Army capabilities. This path holds much promise for future analysis.

2. T-score Analysis

The T-score has been utilized in the TRP as a vehicle for reflecting equal intervals of ability. This system is not critical in providing the commander with a normative evaluation. Simple deciles, calculated from the raw data could suffice for that purpose.[19] The T-score, however, provides a major benefit in the analysis of the resulting data. The most powerful of all statistical tests are the parametric tests. These tests often require the assumption of normality, that is, that the samples be drawn from a normal population. The T-score conversion automatically transforms all data distributions to a normal (50,10) distribution. This allows the application of most parametric tests, as all samples can now be considered to be drawn from a normal distribution. Additional convenience is afforded by the fact that the standard deviation is transformed to a fixed value (10) for all MOEs and this allows several powerful comparison tests to be used.

---

[19]This assumes that the problems addressed in Chapter 4 do not occur, i.e. sparseness of data and synergistic "test problem" effects. See Chapter 4, section A(4).

Thus, parametric tests concerning means can be employed to conduct trend analysis [Ref. 31:pp. 390-396]. The effect of NTC training can thus be verified by the hypothesis testing of MOE means to see if they are increasing (at an appropriate significance level). The differences between MOE means can also be tested to see if some particular combat skills, as measured by the MOE, are being learned better than others. The application of these tests to TRP data (T-scores) can pay big dividends in increased understanding of the combat process, which is beneficial to combat development agencies, and better understanding of the effects of the NTC training on unit performance.

One of the biggest advantages of the use of T-scores, however, is in regression and correlation analysis. As all MOEs are distributed identically after undergoing the T-score transformation, the techniques of multiple regression analysis can be utilized to examine the interrelationship of the MOEs. This analysis could not only include correlation analysis to determine MOE interrelationships, but also could examine the combat process in detail.

Multiple regression could be used to link MOE performance to combat mission outcome. In order to do this, define mission accomplishment as a function that is determined by the values of the MOE T-scores. The particular model represented by the data, i.e. linear, quadratic, etc., could be investigated to determine the true relationship of various MOEs to mission accomplishment.

Thus, by using standard regression techniques on the NTC data base over a sufficient period of time, it would be possible in theory to determine for the first time the empirical relationship of MOEs (or skills) to mission outcome. This sort of study is absolutely impossible under the current ARTEP system for reasons outlined in Chapter 1. This analysis is normally only done by combat developments agencies at great expense. It could now be done _at no additional expense_ to the Army on a routine basis. This analytic effort would _in no way_ affect the training unit or interfere with its operations. Using the TRP as a medium it allows for analysis of the combat process to a degree never before attainable, and could conceivably greatly affect future Army operations and training.

Other techniques, of course, also have much to offer. To investigate sets of MOEs to determine

relationships chi-square contingency tables could be employed. Two MOEs could be contrasted by appropriately subdividing the TRP T-scores into categories and casting a selection of data from the baseline set into contingency tables. These tables could be tested for significance and thereby, MOE relationships examined for their contribution to the mission outcome. As an example of this, "percent planning time forwarded" (PPTF) could be contrasted with "percent friendly personnel survived" (PFPS) to see, if indeed, "haste makes waste". Many other such analyses are possible.

Standard analysis of variance (ANOVA) techniques can be applied to the T-scores to wring even more information out of them. There is no limit to the interactions and relationships that can be investigated between the various MOEs. The T-score does not hold much to offer the unit commander directly, but it contains a wealth of data for the analyst. Yet for all of its usefulness, the T-score is obtained almost painlessly through the TRP structure. Without the TRP, much useful data is absolutely wasted. By employing the TRP, a whole universe of combat developments and operational analysis will be made available for the Army.

3. Nonparametric Analysis

Earlier in this chapter a method was described by which the OC inputs of a subjective nature could be scaled into a "harder" number and employed in a TRP. This system is workable and allows an otherwise "fuzzy" rating scheme to be included in the TRP. The OC rating in itself, however, possesses a great deal of usefulness in its present form. The behavioral sciences have long drawn conclusions from data which is just as subjective. This has been accomplished through the use of nonparametric statistical analysis which is essentially "distribution free" and often not encumbered by the necessary assumption of normality as the parametric tests require. Such statistics are just as fitting in this instance, for the OC is, in essence, doing what many clinical psychologists do in the course of an experiment: he is evaluating human interaction and performance.

The scaling methods of Lindsay [Ref. 30] provide OC data on an interval scale. The interval scale is sufficiently well-defined to permit the application of standard parametric tests in the analysis of data [Ref. 27:pg. 62]. This may not, however, be desirable or feasible once actual data is obtained. An in-depth analysis of

actual OC evaluations would be necessary before statistical inference techniques could be applied. At the time of this writing, the NTC instrumentation has been installed only one month. Therefore, it is impossible to find the actual distribution and reliability of OC input data. Hence, if it is felt that parametric tests are not suitable for this transformed data, nonparametric alternatives are available.

This suggested program is mostly directed toward determining training progress. That is, these nonparametric tests would be most useful in determining if progress has occurred. Two tests are suggested as a basic beginning, and the applicability of other nonparametric tests is limited only by the skill of the analyst.

In turning to nonparametric test theory, it is assumed that the battalion command group/staff evaluation is unchanged from mission to mission. That is, the functioning of the commander and staff is assumed to be a constant process no matter what operation is being conducted. Each mission might require some different specific task to be performed by the unit, but the functioning of the staff element will follow the same basic procedures. If this is the case, then differences in the OC evaluation from the first mission evaluated to the last mission evaluated will

reflect the amount of learning or increase in training proficiency attained by the command/staff element.

In order to assess this, several nonparametric tests for two related samples (i. e. the battalion command group/staff performance at two different times) can be employed. The McNemar test for the significance of changes [Ref. 27:pp. 63-67] can be used for this purpose. In this test each person or item (here, the battalion command group/staff) serves as its own control. The test would take the first evaluation and compare it to a later, subsequent evaluation. In this manner an increased level of training would be shown if the difference in the evaluation ratings received was deemed significant at the chosen level.

An even more powerful test to use in this situation would be the Wilcoxon matched-pairs signed-ranks test. This test almost requires an interval scale, but not quite.[20] The test fully uses all the information available in the data. It is ideally suited to determining the before-and-after effect of events, and could accurately reflect the improvement in the battalion command group/staff operations due to training at the NTC. This test could be employed by

---

[20] See Siegel [Ref. 27:pg. 76] for a complete explanation of test requirements. Siegel discusses an "ordered metric" scale, which falls between the ordinal and interval scale.

using the thirty OC evaluation MOEs as sample items, and the values of the MOEs for the first and subsequent (or last) evaluations as the before-and-after conditions. For a detailed discussion of the actual technique, see Siegel [Ref. 27:pp. 75-83]

Of course, many other nonparametric tests could be applied to the NTC data, both instrumented and OC-input. Such analysis would not only reflect the value of NTC training, but also help to justify future expansion through more incisive cost effectiveness analyses. Nonparametric ANOVA can also be used to examine the relationships of either OC MOEs or instrumented MOEs by determining interactions. It might even be possible, in future training, to structure data for fractional factorial analysis and thus begin an in-depth look at the interactions of the complex combat processes. This information could prove to be extremely valuable to combat developments activities by helping to clarify the key interactions of unit skills on the battlefield.

The NTC is not a test bed, but rather an instrument for accomplishing realistic training. Several statistical methods have been presented here using the TRP structure as a basis. It should be noted that these methods require no

interference with the training unit, and all analysis can be done any time after the completion of an exercise. This system, utilizing the TRP as its heart, allows for exciting, significant analysis of all data without increasing the requirement for additional controllers.

The point to be noted here is that the TRP serves as a vehicle for this analysis, in addition to its already meaningful role as a training evaluation and resource allocation aid. Thus, the TRP has been shown to be a useful device not only to assist in training evaluations, but for combat developments, Army trend analysis and even future budget justification. It is, indeed, a versatile methodology.

VI. CONCLUSIONS AND RECOMMENDATIONS

The Nation, and the Army, have expended a great deal of effort, time, and money in establishing the National Training Center. The realistic training environment at the NTC could serve to greatly increase the Army's combat readiness and improve the chances of survivability in any future conflict. Before these benefits can be realized, however, the NTC must produce a valid and thorough evaluation of unit training performance. It is these evaluations--evaluations of training standards for large units--that assist commanders in identifying unit training weaknesses. Resources can be allocated to training weaknesses and remedial training accomplished only after all shortcomings are clearly identified. The desired result, therefore, is to gain the maximum benefit from limited training resources and produce an Army that performs at a high level of combat proficiency.

An attempt has been made here to examine the current Army training system from its inception to its application at the National Training Center. This examination has focused on the area which presents today's training manager with the most difficulty--the evaluation (and indirectly the

establishment) of training standards for large units. This is a problem of great difficulty, and has not-been properly addressed in recent times.

The creation of the NTC has spurred efforts in this field, and has resulted in an almost revolutionary breakthrough in the training process. The instrumented battlefield will present, for the first time in Army history, a realistic training ground where the synergistic effect of modern Army weapons can truly be evaluated. It will, indeed, provide unit commanders with a statistically reliable, objective training evaluation. Traditional ARTEP feedback processes and AAR's as planned for the NTC do not fully capitalize on this fact.

A. CONCLUSIONS

The Training Readiness Profile attempts to capitalize on the statistically reliable data gathered by the NTC. The NTC is a revolutionary, not evolutionary, step forward. As befits such a system, the TRP is proposed as a methodology for use there. The TRP employs a method previously unused in Army training programs. It is believed that this unique methodology addresses a number of current training problems fully--and takes maximum advantage of the NTC's objective, statistically reliable data. The TRP will allow the NTC to

realize its full potential, and recognizes that the quality of evaluation provided to the training unit is of key importance.

The TRP, by virtue of its structure, accrues the following benefits through its use:

- It provides a simple, reliable, easy-to-understand unit training evaluation. It eliminates "telephone book" sized ARTEP evaluations and dreary statistics.

- It provides an almost wholly objective (machine driven) training evaluation, without possible human subjective bias or prejudice.

- It alleviates the problem of determining large unit subtask standards (a prodigious effort in itself) by utilizing a normative evaluation.

- It allows meaningful evaluation without resorting to grades, and does so in an almost cheat-proof way--by evaluating an entire unit at once.

- It assists commanders, both NTC and unit, in setting standards for scenarios by clarifying issues of performance.

- It literally provides defacto standards, and easily allows for judgmental standards to be inserted.

- It facilitates Army-wide training level standardization by illuminating "Army performance" through the baseline data set, and communicates this "peer performance" to the unit commander. Thus, it tells him what other battalions are capable of.

- It illustrates complementary weaknesses in training and can help to highlight the interactive effect of training deficiencies in combat.

- It provides to the unit commander a solid guide for allocating training resources, by defining (through normative evaluation) the worst areas of unit performance--those that need more training.

- It does not usurp command authority and, indeed, still requires the commander to make allocation decisions and examine his unit's training status in detail.

- It holds great potential for combat developments analysis "as is" without extensive data modification or additional data collection.

- It can be used as a tool for assessing Army-wide readiness by clearly portraying training trends and levels of capability.

- It has a great deal of flexibility inherent in the system and possesses unlimited potential for expansion and further application in several fields.

- It allows the National Training Center to realize its full potential by not only organizing data for unit presentation and use but also by allowing combat developments work to proceed apace simultaneously with training, without mutual interference.

These extensive and valuable benefits are available simply by implementing the Training Readiness Profile at the NTC.

In the preceeding chapters two forms of evaluation have been discussed; the subjective standard evaluation and the objective, normative or relative performance evaluation. Analytically, the normative evaluation works best at the present time. Start-up problems inherent in any complex operation such as the NTC will render arbitrarily fixed standards irrelevant in short order if they are not based on a normative assessment of what is really happening on the instrumented training area. Long term analysis will be necessary to effect the formulation of such fixed standards [Ref. 14:pp. III-9 to III-11]. The NTC neither has this analytic capability at present, nor is it likely to acquire any in the near future as trained operations research analysts (specialty code 49) are not planned or projected

for assignment to the NTC in quantity. An in-depth analysis of NTC operations will be necessary to assess realistic scenario standards, and in the interim (or even permanently) the normative evaluation process is perfectly acceptable. Both normative and fixed standard evaluations may prove useful in the long run with neither supplanting the other.

B. RECOMMENDATIONS

The following recommendations are made for improving the operation of the National Training Center.

1. Implement the Training Readiness Profile

To assist commanders in identifying training weaknesses the TRP concept should be implemented at the NTC. The TRP should be used by the commander with the assistance of the training analysis and feedback (TAF) personnel for directing efforts to determine specific training weaknesses. The relative comparison between MOEs should be used when prioritizing the allocation of training resources. The implementation of the TRP should be a matter of high priority to insure that the maximum training benefits (as outlined above) are availible to all units training at the NTC as soon as possible.

2. Establish NTC/TSM Analysis Cell

Currently, there are no operations research analysts assigned to the NTC section at Ft. Monroe, the agency responsible for the continuing development of the NTC. To provide assistance in the design and growth during the expansion phases of the NTC an operations analysis cell should be assigned to the TRADOC Systems Manager (TSM) office at Ft. Monroe. The analysis cell or group would provide the capability to structure and orient the instrumentation equipment package so as to maximize the benefits of producing unit training evaluations.

3. Establish an NTC Analysis Cell

A group of qualified analysts should be assigned to the NTC to assist with the process of sorting and organizing data obtained from the instrumented battlefield at the NTC. Much of the current analysis applicable to the NTC has been accomplished by operations research analysts assigned to other organizations. This has increased the workloads of these "parent" organizations (such as CATRADA) while introducing almost insurmountable problems of distance and time into all analysis produced. These personnel could also conduct the independent trend analysis necessary to revise the base of MOEs and recommend the updating of standards.

Additionally, the data base at the NTC will prove to be a valuable source of information for many Army agencies. When this data is made available to these organizations, chances are very great that they will want access to the data base and request assistance in retrieving selected sections of the data. Operations research analysts present at the source of this data could prove to be a very cost effective investment for the Army, as it would increase the chances of requestors obtaining the right data the <u>first</u> time it is requested, and it would insure that researchers understand the sources and limitations of all such information. In this way, coordination between the NTC and combat developments agencies would be of a high order.

The TRP program needs acceptance. There is no avoiding the hard fact that realistic combat training evaluation is absolutely necessary if the Army is to fight and win the next war. The TRP is suggested with the goal of attempting to assist the professional unit commander in improving his unit training. It does NOT compare units, but rather helps the unit commander see "how far he has to go". It must be utilized for what it is worth, and accepted in the spirit in which it was meant: as an objective training evaluation.

The implementation of these recommendations will enhance not only the quality of training at the National Training Center (through better feedback), but insure that the instrumented data gleaned by this system is most effectively utilized. In war, a soldier's duty is clear, and the tasks he must undertake are before him. In peace, the soldier can only train for war as he thinks it will be. The realism of that training in peacetime is the key to the soldier's success in wartime. The ultimate in training realism is now before the US Army at the NTC. There is no other facility like it at this time, in the world. It must be employed in such a way as to assure the maximum benefit possible to our troops that train there. To do otherwise, wastes the resources of the Nation.

And costs the lives of her soldiers.

APPENDIX A

LIST OF TRAINING MISSIONS PLANNED FOR THE NTC

The following is a list of tactical missions planned for training units at the National Training Center. These missions fall into two categories: those missions considered to be so important that all training units will execute them while at the NTC, and those other missions which the unit commander can choose to execute or not according to his unit's training plan. Missions identified by an asterisk are considered to be most important.

Movement to Contact *
Hasty Attack *
Deliberate Attack *
Defend in Sector *
Defend from a Battle Area *
Hasty Defense *
Delay in Sector *
Disengagement *
Counterattack *
Defend a Battle Position *
Deliberate Defense
Reconnaissance and Security
Create and Defend a Strongpoint

A series of scenarios of varying intensity will support each of these missions. These packages will be used by the NTC

Operations Group to conduct training and provide orders to training units in a rapid and standardized manner. [Ref. 18:pp. 1-12 to 1-13]

APPENDIX B

DEFINITION OF ELEMENTS UTILIZED IN THE TOP-DOWN ANALYSIS PROCEDURE

The following definitions were utilized in accomplishing the top-down subtask analysis for the National Training Center and are taken from Reference 14, page III-5 and III-6.

1. Missions. A clear, concise statement of the overall purposes of a system.

2. Tasks. (also comprised of any number of subtasks). Objectives which must be accomplished to satisfy the mission.

3. EEAs. Essential Elements of Analysis. These are subcategories of the tasks which lend themselves to analysis. The EEA's should be derived from these tasks to cover all the significant aspects of the force unit's mission. Thus each EEA must relate to one or more tasks.

4. MOEs. Measures of Effectiveness. Quantitative indicators of the general ability of a military force to accomplish an assigned task or mission under a specified set of conditions (must address EEAs).

5. MOPs. Measures of Performance. Quantitative measures of the ability of a particular subunit or weapon system to accomplish a particular task. MOPs may be combined to compute MOEs or may address one or more EEA directly.

6. Data. Facts or statistics that provide descriptive information pertaining to a single event.

APPENDIX C

EXAMINING THE COMBAT PROCESS: THE TOP-DOWN ANALYSIS

The first step taken in the sequence of events was the identification of EEAs. Drawing from ARTEP 71-2 [Ref. 6], and FM 71-2 [Ref. 16], a basic list of combat missions was established, listing in order of importance the combat missions which could be assigned to the heavy task force [Ref. 32:Incl 2]. This prioritized list was then sent to the TRADOC schools to allow them to begin considering MOEs and MOPs to be utilized in evaluating unit training at the NTC. The recommended methodology used in this process was identical to what has already been described, and reflected the hierarchy of Figure 8 [Ref. 32:Incl 3].

Determining the EEA did not prove to be an easy task. The initial NTC Development Plan gave the responsibility for the top-down analysis of combat tasks to the TRADOC combat development community [Ref. 14:pg. III-5]. This task devolved upon the Combined Arms Training Development Activity (CATRADA) at Ft. Leavenworth, who began analysis on the EEA/MOE/MOP structure. The final EEAs were drawn from their logical source: Army doctrine as specified for the

heavy task force in FM 71-2, _The Tank and Mechanized Infantry Battalion Task Force_ [Ref. 16]. In specifying how to organize and operate a heavy task force, FM 71-2 utilized a system approach, and defined the commander's operating systems as:

Maneuver
Intelligence
Air Defense Artillery
Mobility-Counter Mobility
Combat Service Support
Fire Support [Ref. 16:pg. 3-11]

To this was added the commander's interface system of C^3--Command, Control, and Communications. These 7 systems became the EEA for the NTC, to which an eighth system, NBC, was later added. This analysis, taken directly from current Army doctrine, formed the basis from which the MOE/MOP development proceeded.

In the long run these EEA must be considered the best choice. The initial list of EEA proposed bore little resemblance to the hierarchy specified by Figure 8. Those topics as originally proposed were overlapping, or in many cases a function of one another. Some proposed EEA were, in fact, subsets of one another. FM 71-2 states that:

"Although it is true that the actual fighting is done by members of companies, it is the battalion and its commander on which the battle outcome chiefly depends." [Ref. 16:pg. 3-2]

This demonstrates that doctrinally the battalion operations, as directed by the battalion commander, are the heart of the combat process. In fact, the task force is the major application of combined arms fighting ability in warfare today:

> "...the battalion is the level which combines the various arms tactically on the terrain and brings its combined combat powers to bear on the enemy." [Ref. 16:pg. 3-2]

Therefore, any EEA which purport to interactively provide an assessment of mission accomplishment or unit performance must center around the organized application of combat power by the battalion commander. Indeed, this is the primary goal of the NTC:

> "At the NTC the primary target of evaluation and corrective training will be the battalion's ability to orchestrate the application of total combat power." [Ref. 14:pg. I-5]

Thus, it is difficult to conceive of a selection of EEAs that would be meaningful if they did not center around the battalion's operations. Indeed, as these EEA are drawn directly from Army doctrine, any MOE/MOP analysis based upon them must inherently contain the necessary subtasks for a proper training evaluation. As these EEA reflect the battalion's ability to integrate the systems at the disposal of the battalion commander, they will reflect accurately

unit performance once mission accomplishment (outcomes) are determined, for:

> "Tactically skillful commanders have battalions which optimize their effectiveness and minimize their vulnerabilities, and thus habitually execute their missions successfully with minimum losses." [Ref. 16:pg. 3-2]

It is upon this solid framework that the MOE/MOP structure is built. The work done by CATRADA and the service schools in determining the quantitative MOEs and MOPs to be applied to unit training is correctly based on doctrinal underpinnings.

APPENDIX D

THE FOUR TYPES OF MEASUREMENT SCALES

Type of Measurement Scale	Use	Example
Nominal	Identification	Assignment of street numbers, License plates
Ordinal	Ranking without Quantitative Comparison	A is better than B B is better than C
Interval	Ranking on a Non-Absolute Scale	Fahrenheit and Centigrade temperature scales; any measurement scale that provides fixed intervals for comparison but whose zero point is arbitrarily fixed.
Ratio (Cardinal)	Rank Order	Based on an absolute zero. This scale gives order absolute differences, and the true ratio between the quantities measured.

APPENDIX E

IDENTIFYING THE MEASURES OF TRAINING PERFORMANCE

In initially considering the problem of identifying MOE for use at the NTC the above philosophy seems innocuous enough, but it was soon discovered that accurate MOE were difficult to come by. No one had ever dealt with the battalion level combat process to such an introspective degree before. In tackling the problem at CATRADA, a slightly modified MOE development methodology was posited [Ref. 20]. The same hierarchy as demonstrated in Figure 8 was utilized, but in the analysis of the eight EEA efforts were focused on the six interrogatives:

> "The three categories (EEA's,MOE's, and MOP's) can be thought of as answering the questions where, when, what, who, why, and how." [Ref. 20:pg. 25]

and the analysis process followed that of Figure 20. This enabled the CATRADA analysts to compile an initial breakdown of subtasks for the eight EEA which were subsequently organized by the NTC system contractor, Science Applications Incorporated (SAI), of El Paso, Texas, into the first compendium of MOEs/MOPs, entitled <u>National Training Center: EEA's, MOE's, MOP's</u> [Ref. 33]. This represented the "first cut" of ideas regarding what the NTC subtask analysis

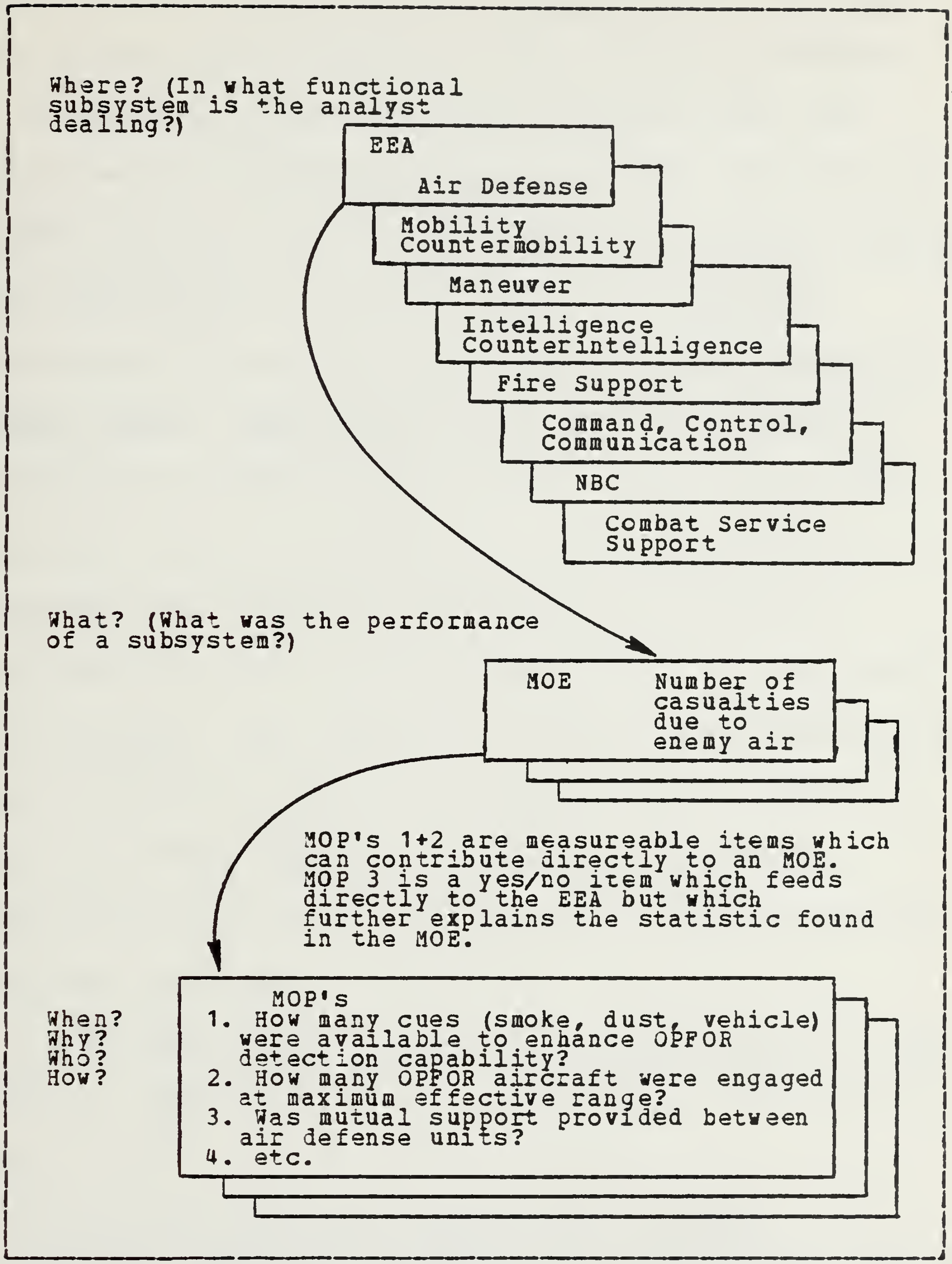


Figure 20: CATRADA MOE Development Methodology

should contain, and were based initially on the experience and judgment of the military officers doing the analysis.

With the additional analytical efforts of SAI, the development of MOE/MOP proceeded apace. The next draft of MOE/MOP contained the first analysis of the original CATRADA work restructured in terms of manual or instrumented information retrieval systems. Entitled Combat Evaluation Program (EEA's, MOE's, MOP's) [Ref. 34], this document related the subtasks of Reference 33 to the level of unit (platoon, company, or battalion), type of information retrieval system (manual or instrumented) and stage of development of the NTC at which the MOE would be introduced to the system. This subtasking process was identical to the method employed by ARI in the structure of the REALTRAIN validation studies [Refs. 8,9,10,11] and represented accepted state-of-the-art analysis of the combat process for the heavy task force.

As SAI began closer integration with the efforts of UTD-CATRADA it became obvious that the subtasking analysis was a mammoth task.

> "In order to address the large number of MOE's characterizing Combined Arms force on force exercises... Systematic analysis is required to avoid "getting lost in the woods" of the combat details. Further, it is necessary for the AAR preparers to have a systematic interpretation procedure so as to provide sufficient comparative data so as to diagnose and guide maximal corrective action. [Ref. 35:pg. 2-2]

The level of detail, if carried to its extremes at the lowest echelons of a unit, threatened to overwhelm any evaluation or feedback system, for combat is simply a very complex, highly synergistic process. Gradually the concept of a "hierarchy of detail" or "fineness of grain" was implemented in the analysis. The analysis was:

> "...organized in a systematic fashion utilizing the concepts of EEA's, MOE's, and MOP's. These quantitative analyses and interpretations provide the why's of the experience: "what happened and when it happened." They provide for fine grain understanding of where "things went wrong" and the desired type of corrective action." [Ref. 35:pg. 2-2]

This concept supported the top-down analysis and enabled SAI and CATRADA to focus the evaluation. Given that the NTC would be structured with limited funds and resources, the entire spectrum of battalion combat operations could not be evaluated to the fine grain desired. Therefore, the continuing analysis for the NTC was structured as follows (see Figure 21):

> "Starting with the core actions at the FEBA, the most important EEA relate to maneuver. This is further supported by each of the concurrent seven EEA areas. As indicated, the analysis will proceed from the core actions out to the supporting EEA's such as Intelligence and Counter-Intelligence." [Ref. 35:pg. 2-3]

The rationale for this procedure was:

> "Detail will focus on the FEBA as represented by BLUEFOR and OPFOR actions and reactions. This represents the primary NTC objective of combat, command, and personnel training in a relative environment.
>
> Timing--Maneuver measures occur in seconds and meters while Intelligence/Counter Intelligence measures occur in hours and kilometers, thus the fine grain detail of combat

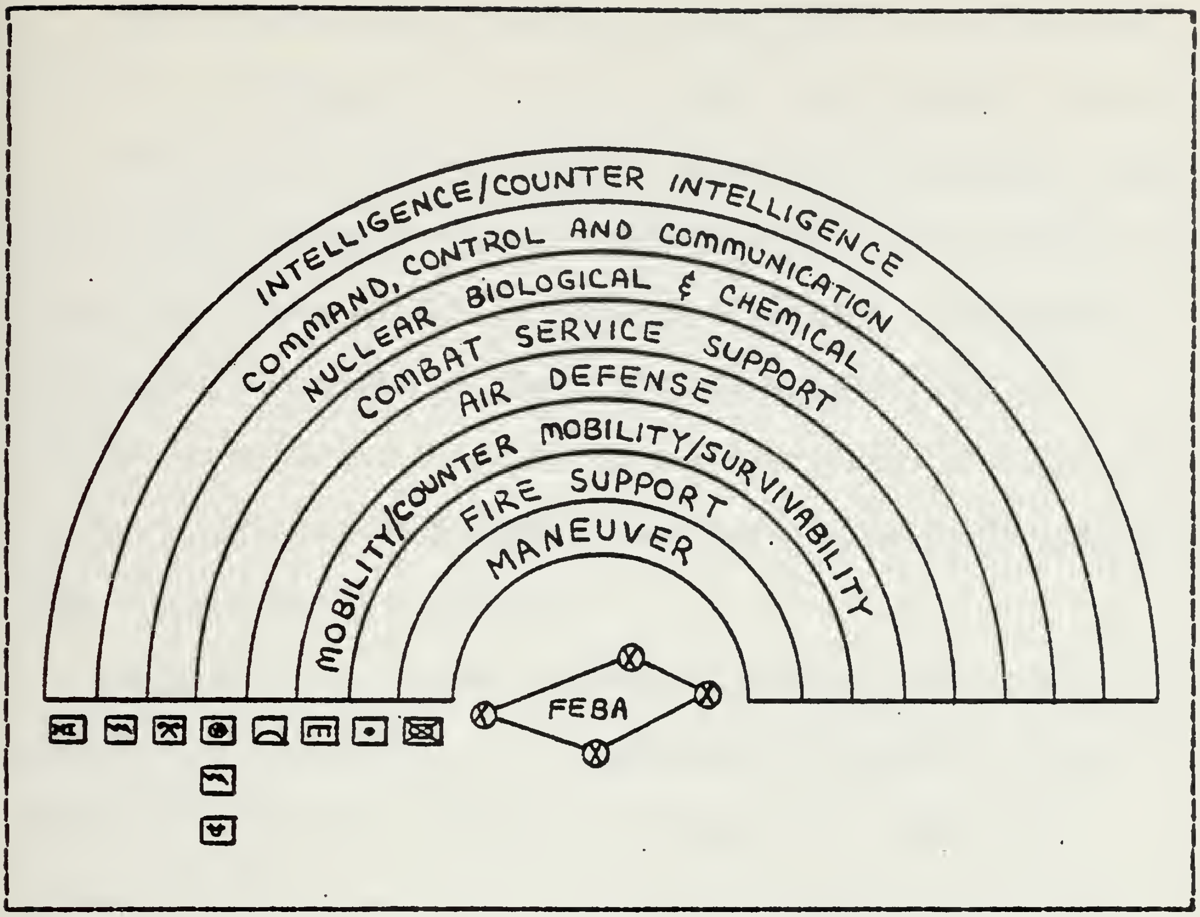


Figure 21: The NTC Combat Evaluational Structure

> critical events of maneuver must be considered prior to the longer term events from an analysis and interpretation viewpoint.
>
> The amount of command impacts in required MOE's are greatest in maneuver and decrease through the subsequent levels of EEA analysis." [Ref. 35:pg. 2-3]

This technique allowed the process of MOE/MOP development and system integration to proceed and culminated in the published Interpretive Guide to the NTC combat evaluation program [Ref. 35] which delineated all MOEs and their subordinate MOPs to be considered at the NTC.

This procedure has resulted in what has to be the most complete indexing to date of the tasks and subtasks involved in battalion task force combat. Under the development plan of the NTC, however, these MOE/MOP are not to be regarded as fixed but rather as initial analysis efforts. Throughout the future operation of the NTC these MOE/MOP will be subject to revision as the training battalions provide more and more information regarding the task force combat process [Ref. 14:pg. III-9 to III-11].

Therefore, the combat evaluation program may rightly be regarded as incomplete. It represents, however, the best efforts of military judgment to dissect the task force combat process. This program compares favorably to most other analyses of this nature [Refs. 36,37]. The analysis is prodigious in its content, and although all aspects of it cannot be quantified under the physical limitations of the NTC, it provides a "base case" analysis and identifies the most critical items to be quantified. These MOE/MOP are truly an accurate state-of-the-art reflection of the combat process of the heavy task force.

APPENDIX F

THE QUANTIFICATION OF TACTICAL PERFORMANCE

The initial MOE/MOP development has followed a lengthy, complicated process and it is recognized that these measures should be reviewed on a recurring basis. Any need for improvements, refinements, and changes will hopefully become apparent through repetitive analysis of the selected MOPs. The importance of these MOPs cannot be overstated since they provide the underlying reason why specific performance criteria are being quantitatively measured by the instrumentation system. The entire instrumented environment at the NTC has been designed for the purpose of recording those specific events and occurrences in the engagement simulation exercise that are necessary to evaluate the training standards for large units--the key problem.

Data collected by the Core Instrumentation Subsystem (CIS) comes from two sources: the instrumented environment which automatically sends objective measures at designated time intervals, and field observer/controllers (FOC's, or later OC's) who input subjective comments and information periodically as the situation dictates.

A. AUTOMATIC DATA FROM THE INSTRUMENTED ENVIRONMENT

This environment consists of a complex interface that combines voice and video recording with other telemetry to provide certain measured data to the CIS. This data is manipulated and filed so that it represents an accurate portrayal of the combat activities that transpired during an exercise. The actual quantifiable items being measured are outlined below.

1. Position Location

Every instrumented player will have his location measured in UTM coordinates, using metric system units, each 30 seconds. If he has moved less than 16 meters since the previous update the move is considered insignificant and is not recorded. The measured data will be filed by time sequence and individual player item identification [Ref. 18:pg. 3-14]. As an aid to map coordinates there are six video teams that will be assigned to various sectors of the battlefield. Film of actual locations and surrounding terrain will provide enhanced insight into actual field positions and will reflect the use of available terrain features for cover and concealment. Knowing the exact locations of players allows movement distances and ranges between various players to be accurately computed.

2. <u>Weapons Firing/Effects Event</u>

When a weapon is fired by an instrumented player a series of information is automatically recorded: time of firing, identification of firer, and his location. When a simulated round impacts on or near an instrumented player similar information is recorded: time of impact, identification of the target player, and his location. As these events are recorded by the CIS a time coincidence algorithm is activated which attempts to establish a pairing between a firer and a target. If a weapons pairing is satisfactorily confirmed the recorded information is updated to reflect the pairing action and the result of the engagement i.e. near miss, hit, or kill. Since player identifications are assigned prior to exercise commencement the type of weapon system firing and the type of target hit are known from these events. (For a more detailed discussion of the pairing process refer to section 3.2.2.2, Ref. 18:pg. 3-15.) From these recorded events numerous data items can be compiled. The number of rounds of ammunition fired by individual weapon system type, the number of enemy targets of various types hit and/or killed, engagement distances between firers and targets, the current status of each force (i.e. number of players alive by weapon type),

and many other such objective data values are easily calculated. Existing instrumentation at the NTC is programmed for these and other calculations.

3. Communication Keying Event

When an instrumented radio set transmits a message the beginning and ending time of the transmission is automatically recorded. This allows for the computing of the number and duration of transmissions by radio set [Ref. 18:pg. 3-18]. Additionally, selected radio nets will be recorded and/or monitored to allow for analysis of radio message content if desired. This will be helpful in disclosing communication security (COMSEC) violations or evaluating operators for proper radio procedure.

B. DATA FROM OBSERVER/CONTROLLERS (OC'S)

Observer/Controllers (OC's) will be detailed to each unit down to the platoon level for the purpose of observing all aspects of a heavy battalion's combat operations. They will use ARTEP 71-2 for performance comparison when conducting evaluations or providing information to the CIS or subsequent feedback to the evaluated unit.

> "The purpose of the ...OC...is to provide timely, on-the-scene information (made as accurate as possible through controls) which cannot be obtained automatically." [Ref. 38:pg. 4-2].

Certain actions within the various levels of activity are not suitable for direct instrumented collection. For example, no instrumentation can indicate whether or not the commander used available battlefield intelligence in developing his plan of operations. The OC's will serve as the collection source for this type of unquantifiable data [Ref. 38:pg. 1-12].

Prior to commencement of an exercise phase each OC will be given a selected list of elements of information (EI) which he must provide to the CIS. These EI come from a developed listing of approximately 200 items that cover such areas as target acquisition, manuever, fire support, command and control, logistics, administration, and others. The OC will conduct his subjective evaluation based on current doctrinal publications and transmit his report to the CIS. The evaluations can be in the form of a nominal measure ranging from 0 to 9 or alternately a free format message describing the situation [Ref. 18:pp. 3-25,3-26].

A good example of the OC and CIS interface is provided by the use of indirect fire. When the CIS is notified that indirect fire is being aimed at a unit the computer provides time, location of impact, number and type of rounds to be fired and the expected casualties produced based on the

recorded information about position locations. This information is relayed to the OC who marks the fires on the ground and then uses his own judgement and expertise to determine, if in fact the expected casualties did occur. The OC then provides the CIS with actual combat casualties and takes appropriate action to designate the killed players by deactivating their MILES equipment with his controller key device. It is obvious that the OC input has a direct influence on data collected in this type of engagement. All OC input is tagged so if a particular individual is considered to have rendered improper assessments a correction is possible.

In developing an objective training evaluation the use of subjective inputs from OC's is highly questionable. The affect of indirect fire casualty assessments and some other activities, however, must necessarily be incorporated into the battle outcome data as no other system exists to simulate these combat effects. Thus, in the scheme of evaluating training standards, the OC's will play an important, albeit less statistically reliable, role.

APPENDIX G

THE GENERATION OF STATISTICAL EVALUATIONS

Presented in the previous section is the actual data being collected by the instrumentation at the NTC. The list of items measured appears to span only a limited array of the activities planned for the NTC exercises. However, a thorough, imaginative eye will note that the data being collected provides the information necessary to conduct an extensive objective training evaluation. These collected data items can be manipulated and aggregated to generate useful information in many different forms. This appendix will be devoted to explaining how these data items can be utilized to provide an objective evaluation of unit tactical performance for the training exercises at the NTC.

A. GENERAL PROCEDURE

At the beginning of an exercise each instrumented player is matched with a designated identification code. Any data that is collected concerning a player will be stored in his unique player history file (PHF); this includes weapon firings, movement, change in status from alive to killed, etc. When an event occurs the data describing that event is

recorded and stored in the appropriate PHFs. To facilitate the aggregation of data by units (platoons, companies, etc.) each player is assigned to a specific tactical unit. A listing of identification codes for all members belonging to a unit is maintained in the CIS for reference and for sorting data. During an exercise, players may be cross-attached to adjacent elements. When this occurs all subsequent activities concerning that player will be credited to the new unit.

Since participating units in the NTC exercises are comprised of many diverse weapon system types, i.e. dismounted infantry, tanks, various types of anti-tank weapons, mortars, artillery, etc., it is desirable to have a method of comparing the "combat power" capability of two opposing units. Military planners and operations analysts, for at least thirty years, have used various "firepower-score" approaches for aggregating the diverse combat capabilities of a heterogeneous military force into a single number that measures combat power. Each weapon type is assigned a score or weighted effectiveness index (WEI) that represents its combat potential and the combat power of a unit is obtained by summing the scores of all live players belonging to that unit and forming the weighted unit value

(WUV) [Ref. 39:pp. 85-86]. TRADOC has provided a listing of numbers referred to as the WEI/WUV table to the NTC (see Figure 22), and this is stored in the CIS master file to supply the firepower scores necessary for computing the combat power of units.

The NTC instrumentation is currently programmed to compute and update selected statistics every five minutes for a maximum of 50 units. Each platoon, company, and battalion is considered a separate unit. For example, a battalion consisting of only 3 companies with 4 platoons in each company would be a total of 16 units; i.e. 12 platoons, 3 companies, and 1 battalion. The statistics are aggregated upward through every echelon. The instrumentation is flexible and can accomodate reorganization of entire units. All statistics computed following a change in unit organization will incorporate those players who currently belong to the designated unit. The data, and subsequent computed statistics, are compiled for each exercise segment. Exercise segments are event driven and can last from 4 to 48 hours with an average duration of about 8 to 12 hours. The segments are delineated by: natural engagement breaks; a change in tactical mission; moving to a new terrain area; a change in environmental conditions; or a command decision.

These are the weighted values used at the NTC, as derived from TRADOC sources.

blue player type:		BLUE 1	RED 2	red player type:	
tank	1	60.0	67.8	1	tank
APC	2	10.0	7.5	2	BMP
APC w/TOW	3	37.0	10.7	3	BMP w/ M-60
manpack	4	0	32.93	4	BMP w/sagger
manpack w/viper	5	2.22	8.9	5	BRDM
manpack w/dragon	6	23.68	25.2	6	BRDM w/sagger
manpack w/M-16	7	1.1	24.72	7	ZSU 23-24
manpack w/M-60	8	2.28	0	8	howitzer
redeye	9	0	0	9	manpack
GSR	10	0	18.5	10	manpack w/sagr
jammer	11	0	1.0	11	manpack w/M-16
collector	12	0	2.11	12	manpack w/M-60
truck	13	0	0	13	jammer
ADA	14	0	0	14	collector
mortar	15	0	0	15	truck
helicopter	16	0	0	16	ADA
-------	17	0	0	17	mortar
-------	18	0	0	18	helicopter

Figure 22: Weights for Computing Unit Scores

Since each segment presents a different situation to the training unit it seems appropriate to consider each segment as a separate evaluation of the unit performance [Ref. 18:pp. 3-11 to 3-27].

All recorded data items are objective numbers that can be used in various mathematical operations to generate a truly objective measure of unit performance. The MOEs/MOPs developed by CATRADA provide the basis for constructing a unit evaluation. If the MOEs/MOPs were to change in the future, then the recorded data would be altered to a different format to provide a new objective measure. The NTC instrumentation is programmed for expansion in phases over the next few years. As the instrumented system grows the number and variety of statistics available for computation can be increased accordingly.

B. THE STATISTICAL REPRESENTATION OF MOES

The ultimate purpose of the NTC instrumentation is to record data for use in providing a training evaluation to the heavy task force. To achieve this end, the MOE/MOP structure developed jointly by CATRADA and SAI had to be translated into calculable statistical entities. This difficult process was conducted over a long period of time on an ad hoc basis[21] between the two agencies, and resulted in the player tactical performance kernel statistics

---

[21] Per conversations with Rich Scagglioni, SAI La-Jolla, CA, and Norma Perez, SAI-El Paso, TX, in November and December of 1981. This interfacing process between the contractor, SAI, and the developing agency, here CATRADA, was not documented in writing.

[Ref. 40:pg. 3-22 to 3-26]. The major areas addressed by the statistics include: overall engagement effectiveness, movement, weapon effectiveness, and communications. These statistics provide quantitative measures of unit performance in each of the specified areas, and currently serve as the objective base that supports the training evaluation provided to each unit.

1. Overall Engagement Effectiveness

The statistics generated in this category provide a general overview of the unit's ability to accomplish its assigned mission. The following numbers will be computed each update period for both the friendly and the OPFOR units.

Number of casualties-- A count of the number of instrumented players killed.

Overall force value--The sum of the combat power for live players belonging to that unit. The combat power is obtained from the WEI/WUV table stored in the CIS.

Overall force value loss measure--Total combat power lost by the unit during that update period.

Force in contact measure--When a pairing event between a firer and a target is recorded, the platoons to which these players are attached is considered to be in contact. The measure will be the total number of platoons in contact (each platoon will only be counted once during an update period).

Force engagement measure--When a pairing event occurs, the firer weapon will be considered engaged. The total force engaged on each side will be the sum of combat power engaged (each player counted only once for an update period). [Ref. 40:pp. 3-22 to 3-23]

2. Movement

The following statistics represent the unit's ability to maneuver on the battlefield. These statistics can be used to provide an indication of how well the unit used available terrain, whether or not an adequate route reconnaisance was performed, etc. Alone, these numbers are not very useful in conducting a training evaluation, but when coupled with other information, such as overall engagement effectiveness, they provide an insight to possible deficiencies. Each number will be computed for both the OPFOR and friendly units for each update period.

Rate of advance--The center of mass of each unit will be computed based on the locations of players belonging to the unit. The distance between beginning and ending center of mass points divided by the length of time between updates will produce this average rate of advance value.

Rate of advance toward a destination--The distance of unit movement projected on the line from the unit's center of mass to a manually designated objective divided by the length of time. [Ref. 40:pg. 3-23]

3. <u>Weapon Effectiveness</u>

The ability of soldiers to properly and efficiently use their weapon firepower in combat is a vital necessity for survival. However, being effective with weapons does not ensure success on the battlefield; success is achieved through the proper integration and utilization of several combat skills. The statistics in this area do provide an effectiveness measure for weapon employment. They can be used to evaluate marksmanship capability, to determine if soldiers are estimating distances correctly before directing fire at a target, etc. Numbers of this type provide measures that can be used to evaluate key areas of overall unit performance. Every number will be computed each update period.

<u>Rounds fired</u>--The number of rounds fired by each instrumented weapon type for both friendly and OPFOR units. A cumulative number of rounds fired by each friendly unit for tank main gun and coax, TOW, Dragon, and Viper. A similar cumulative count for OPFOR Sagger, tank main gun and coax. One firing message of the coax will be considered 100 rounds.

<u>Firing results</u>--The number of near misses, hits, and kills by friendly units with tank main gun and coax, TOW, Dragon, and Viper against OPFOR targets of tank, BMP and BRDM-2. The associated cumulative values for the engagement segment will be maintained.

Ratio of firing results--The ratio of total firings to near misses, hits, and kills by friendly unit for tank main gun and coax, TOW, Dragon, and Viper and a similar ratio for Sagger, tank main gun and coax OPFOR weapons.

Weapon fractional kill effectiveness--The total value (from the WEI/WUV table) of opposing players killed by tank main gun and antitank weapons (TOW, Dragon, and Viper for friendly and Sagger for OPFOR) divided by the total value of opposing players killed.

Weapon engagement range--The range from a weapon to the target in a pairing event, by weapon effect (near miss, hit, kill) for tank main gun and antitank weapons (TOW for friendly and Sagger for OPFOR). A cumulative frequency count will be collected in 15 range intervals of 200 meters each.

Mean kill range--The average distance between friendly and OPFOR tank main guns and antitank weapons (TOW and Sagger) when a vehicle kill was obtained. [Ref. 40:pp. 3-23 to 3-24]

4. Communications

A unit's ability to effectively employ radio communications has an important impact on battle outcome. The following statistics can be used to support and identify suspected training deficiencies reported in the unit evaluation.

Number of communications--The number of radio transmissions initiated during the update period.

Average transmission duration--Average length of all radio transmissions for that period as well as the average duration of transmissions during the entire engagement segment. These average durations will be computed for each unit as well as aggregated to an overall BLUEFOR and OPFOR average duration.

Significant transmissions--Total number of transmissions exceeding 25 seconds, but less than 55 seconds. Total number of transmissions lasting 55 seconds or longer. [Ref. 40:pp. 3-24 to 3-25]

These statistics represent the final culmination of the Combat Evaluation Program initiated at CATRADA to support the NTC. The exhaustive top-down analysis of the heavy task force combat process yielded a series of EEA, MOE, and MOP which were translated into computer hardware and software by SAI, and the resulting structure of evaluation as it exists today at the NTC has been presented here.

APPENDIX H

CURRENT LIMITATIONS OF QUANTIFICATION

Every operation, no matter how detailed or complex, is designed to perform specific functions. With meticulous effort devoted to operational design numerous functions are possible, but there will always be some cut-off point or limit where the effort must end. The NTC instrumented environment is an example of such a complex operation. Following the extensive MOE development, considerable effort was directed toward the evaluation process itself. Since the planners were restricted to using "off-the-shelf" instrumentation and were held to a relatively short preparation time, initial work has been devoted to quantifying those data sources deemed most crucial to assisting a heavy battalion task force in identifying training deficiencies.

The paramount objective of the NTC is to provide a realistic environment to facilitate combat learning and evaluate a unit's ability to survive and accomplish its mission. The NTC instrumentation has been oriented toward collecting data in support of the combat activities that

occur within the span of the participating soldiers, and it does this exteremely well. Real time casualty assessment with MILES and other core instrument systems provide a fairly thorough coverage of the critical tactical events of the engagement simulation exercises. The quantitative measures, as outlined earlier, provide the information necessary to reconstruct "what, when, and where" for the analysis of tactical events. This represents a major accomplishment over the current system of subjective training evaluations. However, there are aspects of the combat process that are neglected. These areas are the combat activities that do not impact directly on the learning of the participating soldiers. A discussion of these areas is provided only to illustrate the limitations of the NTC.

The EEAs for the NTC have been displayed pictorially as shown in Figure 21. One can see from this Figure that the EEA form concurrent rings that emanate outward from the forward line of troops (formerly refered to as the FEBA). This is to indicate that the crucial EEA which impact most on a battalion task force evaluation begin at the site of the battle. When a friendly unit maneuvers against a thinking OPFOR capable of counter moves a learning process

results. Activities occurring further away from the battle site are important in the sequence of combat learning, but the impact is not immediately felt by the participating soldiers. These more distant EEA require additional time and examination to determine their impact on the battle outcome and thereby produce any degree of benefit to combat learning. Thus, they are less crucial to analysis as performed at the NTC.

Since the quantification of combat activities at the NTC had to be somewhat restricted, only the most crucial areas could be covered by instrumentation in the initial phases of NTC development. As very aptly put,

> "Maneuver measures occur in seconds and meters while Intelligence / Counter Intelligence measures occur in hours and kilometers, thus the fine grain detail of combat critical events must be considered prior to the longer term events from an analysis and interpretation viewpoint." [Ref. 35:pg. 2-3]

Using this rationale, the initial instrumentation at the NTC has been directed toward collecting data in support of the "close-in" EEAs which include: maneuver, fire support, mobility/counter mobility/survivability, and air defense. The major limitations of the NTC instrumentation therefore fall under the headings of the four outer most EEAs of Figure 21, which are: combat service support, nuclear,

biological, and chemical (NBC), command, control, and communications (C^3), and intelligence/counter intelligence.

The limitation of the instruments to provide specific quantifiable measures to support the evaluation of these outer EEAs is well recognized. Field observer/controllers (OC) are directed toward gathering information and data needed to evaluate these areas. The training analysis and feedback (TAF) personnel monitor all collected data and then search for the cause of identified training deficiencies, in an attempt to relate this essentially unquantifiable data to the mission outcome.

Thus, in any further discussion of the evaluation of training standards, an important point to consider will be what impact these limitations have on the task force evaluation. The purpose of the NTC, once again, is to provide a realistic combat environment. The instrumentation package can now accurately display the tactical combat events which occur during engagement simulation exercises. The TAF and OC personnel will assist the unit commander and his staff in identifying what they suspect as training deficiencies, but it is ultimately the responsibility of each commander and his staff to determine the cause of those weaknesses. With the after action reviews and diagnostic

take-home package provided to the unit by the NTC, it should be able to conduct remedial training at its home station and identify the appropriate deficiency that caused a particular problem, without the use of NTC instrumentation.

These system limitations have been recognized and efforts are being made to reduce them. The NTC is being expanded in phases over the next few years. As more time and money are made available the instrumentation capabilities will become more sophisticated, and a broader spectrum of data will be collected. Much effort has been devoted to analyzing the complexity of the combat process. When these details can be accurately developed to represent the soldier's actions then more specific MOEs/MOPs can be selected for measuring by instruments. Finally, the EEAs in the outer rings of Figure 15 are not completely quantifiable; the data necessary to measure the performance in these areas is not well defined. Until the "fuzzy area" of what exactly is intelligence/counter intelligence or C^3, (and how they should be evaluated) is established, these areas cannot be accurately quantified.

The major point to be made is that the NTC instrumented environment possesses certain limitations. The instrumentation is collecting data to support EEA deemed

most crucial for evaluation of tactical combat events near the site of the battle, and this it does well. The EEA that represent longer term activities are not currently adequately quantifiable. There are sources at the NTC and at home stations to assist in evaluating unit performance on these EEA; the realistic battlefield at NTC is not crucial in producing an evaluation of them. Limitations of the NTC instrumented environment are recognized and during the expansion phases of NTC some of these limitations will be eliminated.

LIST OF REFERENCES

1. Department of the Army, FM 100-5, Operations, 1 July 1976.

2. Department of the Army, TC 21-5-1, Training Management Digest, Training Management: An Overview, No. 1, April 1973.

3. Department of the Army, TC 21-5-2, Training Management Digest, Performance-Oriented Training, No. 2, June 1974.

4. Department of the Army, FM 21-6, How to Prepare and Conduct Military Training, 3 November 1975.

5. Department of the Army, TC 21-5-7, Training Management in Battalions, 31 July 1977.

6. Department of the Army, Army Training and Evaluation Program 71-2, Army Training and Evaluation Program for Mechanized Infantry/Tank Task Force, 17 June 1977.

7. Department of the Army, FM 17-19E1/2 Soldier's Manual-Armor Crewman, 8 August 1979.

8. U.S. Army Research Institute for the Behaviorial and Social Sciences Research Report 1203, REALTRAIN Validation for Rifle Squads II: Tactical Performance, by L.L. Meliza, T.D. Scott, R.I. Epstein. March 1979.

9. U.S. Army Research Institute for the Behaviorial and Social Sciences Research Report 1204, REALTRAIN Validation for Armor/Anti-Armor Team, by T.D. Scott, L.L. Meliza, G.D. Hardy, J.H. Banks, and L.E. Word, 30 March 1979.

10. U.S. Army Research Institute for the Behaviorial and Social Sciences Research Report 1213, REALTRAIN Validation for Rifle Squads III: Tactical Performance During Movement to Contact, by T.D. Scott, J.H. Banks, G.D. Hardy, and R.H. Sulzen, July 1979.

11. U.S. Army Research Institute for the Behaviorial and Social Sciences Research Report 1218, Armor/Anti-Armor Team Tactical Performance, by T.D. Scott, L.L. Meliza, G.D. Hardy, J.H. Banks, July 1979.

12. U.S. Army Training and Doctrine Command Paper, National Training Centers, 23 May 1977.

13. Department of the Army, Army Regulation No. 350-50, National Training Center (NTC), 15 March 1980.

14. U.S. Army Training and Doctrine Command Paper, National Training Center Development Plan, 3 April 1979.

15. Department if the Army, FM 71-1, The Tank and Mechanized Infantry Company Team, 30 June 1977.

16. Department of the Army, FM 71-2, The Tank and Mechanized Infantry Battalion Task Force, 30 June 1977.

17. Department of the Army, FM 71-100, Armored and Mechanized Division Operations, 29 Sept 1978.

18. AMEX Systems Incorporated Report, Critical Item Development Specification for the Core Instrumentation Subsystem (CIS) Software for the Phase I National Training Center Instrumentation System (NTC-IS), 15 July 1981.

19. U.S. Army Training and Doctrine Command Pamphlet 11-8, (Draft) Cost and Operational Effectiveness Analysis Handbook, 15 November 1974.

20. U.S. Army Combined Arms Training Development Activity (UTD-NTC) Working Paper, MOE Development Methodology, by D. Chase and J. Ireland, undated.

21. U.S. Army Combat Developments Command Pamphlet 71-1, Force Developments-The Measurement of Effectiveness, January 1973.

22. DCS for Operations, Readiness, and Intelligence, US Army Training and Doctrine Command, Threat Monograph: Electronic Warfare, A Battalion Success Story, December 1976.

23. Thorndike, R.L.,ed., Educational Measurement, 2nd Edition, American Council on Education, 1971.

24. Commander, National Training Center Letter to Deputy Commander, Combined Arms Training Development Activity, Subject: Scenario and Teaching Point Development, 20 March 1980.

25. Dixon, W.J. and Massey, F.J., Introduction to Statistical Analysis, 3rd Edition, McGraw-Hill Inc., 1969.

26. The Psychological Corporation Bulletin, Test Service Bulletin No. 48, January, 1955.

27. Siegel, S., Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill Inc., 1956.

28. U.S. Army Combined Arms Training Development Activity (UTD-NTC) Concept Paper, Battalion Command Group and Staff Evaluation and Feedback at the National Training Center, by J.S. Williams and P. Rock, undated.

29. U.S. Army Combined Arms Training Development Activity Letter to Commander, TRADOC and Commander, NTC, Subject: Battalion Command Group Observation Plan for NTC Engagement Simulation, 28 April 1981.

30. Lindsay G.F., On Constructing Interval Scales Using Data Resulting form Categorical Judgements, classroom instructional handout, Naval Postgraduate School, September 1981.

31. Freund, J.E., and Walpole, R.E., Mathematical Statistics, 3rd Edition, Prentice-Hall Inc.,1980.

32. U.S. Army Combined Arms Center Letter to TRADOC Schools, Subject: Combat Mission Analysis for the National Training Center (NTC), 7 Feb 1979.

33. Science Applications Incorporated Internal Staff Paper, National Training Center: EEA's, MOE's, MOP's, undated.

34. Science Applications Incorporated Report, Combat Evaluation Program, (EEA's,MOE's, and MOP's), originated by USA CATRADA, by D. Hansen and G. Sikich, undated.

35. Science Applications Incorporated Report, us;NTC Combat Evaluation EEA, MOE, MOP Interpretative Guide, by D. Hansen and G. Sikich, 30 November 1979.

36. Science Applications Incorporated Technical Report, A Cross Indexing Structure for DARPA, USA NTC, USA ARTEP, and MCCRES Measures of Combat Effectiveness, Vol.I, by D. Hansen, S. Kravitz and G. Sikich, 18 September 1980.

37. Science Applications Incorporated Report, Cross Index and Keywords for MOE's Derived from DARPA, NTC, ARTEP, MCCRES Sources, Vol.II, by D. Hansen, S. Kravitz and G. Sikich, 15 August 1980.

38. U.S. Army Combined Arms Training Development Activity (UTD-NTC) Coordinating Draft, Field Observer-Controller's Handbook (NTC), 1 July 1981.

39. Taylor, J.G., Force-on-Force Attrition Modelling, Operations Research Society of America, 1981.

40. Science Applications Incorporated Report, NTC-1221-05, Requirements Design Specification for the Core Instrumentation Subsystem Software (125 Players-Expandable to 500 Players), with Change 12, 16 October 1981.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 55 Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
4.	Chief TRADOC Research Element Monterey Naval Postgraduate School Monterey, California 93940	1
5.	Associate Professor James K. Hartman Code 55Hh Department of Operations Research Naval Postgraduate School Monterey, California 93940	1
6.	Associate Professor S. H. Parry Code 55Py Department of Operations Research Naval Postgraduate School Monterey, California 93940	5
7.	Headquarters US Army Training and Doctrine Command ATTN: Director, Studies and Analysis Directorate, Mr. S. Goldberg Fort Monroe, Virginia 23651	1
8.	Associate Professor A. L. Schoenstadt Code 53Zh Department of Mathematics Naval Postgraduate School Monterey, California 93940	1
9.	Professor James G. Taylor Code 55Tw Department of Operations Research Naval Postgraduate School Monterey, California 93940	1

10. Office of the Commanding General 1
US Army Training and Doctrine Command
ATTN: General Glenn Otis
Fort Monroe, Virginia 23651

11. Headquarters 1
US Army Training and Doctrine Command
ATTN: ATCG-T (BG Morelli)
Fort Monroe, Virginia 23651

12. Office of the Commanding General 1
US Readiness Command
ATTN: General Donn A Starry
MacDill AFB, Florida 33621

Deputy Under Secretary of the Army 1
for Operations Research
Room 2E261, The Pentagon
Washington, D. C. 20310

14. LTG Howard Stone 1
Commanding General
US Army Combined Arms Center
Fort Leavenworth, Kansas 66027

15. Director 1
Combined Arms Combat Development Activity
ATTN: ATZL-CAC-A (Mr. Lee Pleger)
Fort Leavenworth, Kansas 66027

16. Director 1
Combat Analysis Office
ATTN: Mr. Kent Pickett
US Army Combined Arms Center
Fort Leavenworth, Kansas 66027

17. Command and General Staff College 1
ATTN: Education Advisor
Room 123, Bell Hall
Fort Leavenworth, Kansas 66027

18. Dr. Wilbur Payne, Director 1
US Army TRADOC Systems Analysis Activity
White Sands Missile Range, New Mexico 88002

19. Headquarters, Department of the Army 1
Office of the Deputy Chief of Staff for
Operations and Plans
ATTN: DAMO-2D
Washington, D.C. 20310

20. Commander 1
US Army Concepts and Analysis Agency
ATTN: MOCA-WG
8120 Woodmont Avenue
Bethesda, Maryland 20014

21. Director 1
US Army Material Systems Analysis Activity
ATTN: DRXSY-CM (Mr. Bill Niemeyer)
Aberdeen Proving Grounds, Maryland 21005

22. Director 1
Combined Arms Training Development
Activity
ATTN: ATZLCA-DS
Fort Leavenworth, Kansas 66027

23. Director 1
USATRASANA
ATTN: Mr. Ray Heath
White Sands Missile Range, New Mexico 88002

24. Director 1
Combat Developments, Studies Division
ATTN: MAJ W. Scott Wallace
US Army Armor Agency
Fort Knox, Kentucky 40121

25. Commandant 1
US Army Field Artillery School
ATTN: ATSA-CD-DSWS
Fort Sill, Oklahoma 73503

26. Director 1
Combat Developments
US Army Aviation Agency
Fort Rucker, Alabama 36362

27. Director 1
Combat Developments
US Army Infantry School
Fort Benning, Georgia 31905

28. Director 1
Combat Developments
ATTN: ATZA-CDE (CPT James Mudd)
US Army Engineer School
Fort Belvoir, Virginia 22060

29. Director 1
Combat Developments
ATTN: ATSA-CDF-S
US Army Air Defense Agency
Fort Bliss, Texas 79905

30. Commander 1
US Army Logistics Center
ATTN: ATCL-OS (Mr. Cammeron/CPT McGrann)
Fort Lee, Virginia 23801

31. Commandant 1
US Army Signal School
Combat Developments
ATTN: LTC Harnagal
Fort Gordon, Georgia 30905

32. MAJ Jeffrey L. Ellis, USA 5
Department of Operations Research, Code 55EI
Naval Postgraduate School
Monterey, California 93940

33. CPT Richard L. Wampler 1
HQ, TRADOC ODCST
Fort Monroe, Virginia 23651

34. CPT John S. Furman 1
150 E. Johnson Street
Fond du Lac, Wisconsin 54935

35. Commander 2
US Army Training and Doctrine Command
ATTN: ATTG-C
Fort Monroe, Virginia 23651

36. Director 2
Unit Training Directorate
US Army Combined Arms Training Developments
Activity
ATTN: ATZL-TDD-N (LTC Northrop)
Fort Leavenworth, Kansas 66027

37. Commander 1
National Training Center
Fort Irwin, California 92311

38. Commander 1
National Training Center
ATTN: AFZJ-OG (COL Darling)
Fort Irwin, California 92311

39. Commander 1
National Training Center
ATTN: AFZJ-OG (LTC Bezner)
Fort Irwin, California 92311

40. Director 1
Training Developments
ATTN: ATZK-TD
US Army Armor Agency
Fort Knox, Kentucky 40121

41. Commandant 1
US Army Field Artillery School
Training Developments
ATTN: ATSA-TD
Fort Sill, Oklahoma 73503

42. Director 1
Training Developments
US Army Infantry School
Fort Benning, Georgia 31905

43. Director 1
Training Developments
US Army Engineer School
Fort Belvoir, Virginia 22060

44. Director 1
Training Developments
US Army Air Defense Agency
Fort Bliss, Texas 79905

45. Commandant 1
US Army Signal School
Training Developments
Fort Gordon, Georgia 30905

46. Commander 2
US Army Forces Command
DCS Operations
ATTN: AFIN-CS
Fort McPherson, Georgia 30330

47. Headquarters 1
US Army Forces Command
Chief of Staff
ATTN: AFCS
Fort McPherson, Georgia 30330

Thesis
F95 Furman
c.1

A methodology for the evaluation of unit tactical proficiency at the National Training Center.